Experiment Design for Systems Biology

by

Joshua Farley Apgar

B.A. Johns Hopkins University (1999)

Submitted to the Department of Biological Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2009

Author .....
Department of Biological Engineering
August 6th, 2009

Certified by .....
Bruce Tidor
Professor of Biological Engineering and Computer Science
Thesis Supervisor

Certified by .....
Forest M. White
Associate Professor of Biological Engineering
Thesis Supervisor

Accepted by .....
Alan J. Grodzinsky
Chair, Graduate Program Committee

This Doctoral Thesis has been examined by the following Thesis Committee:

Examined by .....

Bruce Tidor, Ph.D.
Committee Member and Thesis Supervisor
Professor of Biological Engineering and Computer Science
Massachusetts Institute of Technology

Examined by .....

Karl Dane Wittrup Ph.D.
Committee Chair
JR Mares Professor of Chemical Engineering and Bioengineering,
Massachusetts Institute of Technology

Examined by .....

Forest M. White, Ph.D.
Committee Member and Thesis Supervisor
Associate Professor of Biological Engineering
Massachusetts Institute of Technology

Examined by .....

Jacob K. White
Committee Member
Cecil H Green Professor of Electrical Engineering and Computer Science
Massachusetts Institute of Technology

Experiment Design for Systems Biology

by

Joshua Farley Apgar

Submitted to the Department of Biological Engineering
on August 6th, 2009, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

Mechanism-based chemical kinetic models are increasingly being used to describe biological signaling. Such models serve to encapsulate current understanding of pathways and to enable insight into complex biological processes. Despite the growing interest in these models, a number of challenges frustrate the construction of high-quality models.

First, the chemical reactions that control biochemical processes are only partially known, and multiple, mechanistically distinct models often fit all of the available data and known chemistry. We address this by providing methods for designing dynamic stimuli that can distinguish among models with different reaction mechanisms in stimulus–response experiments. We evaluated our method on models of antibody–ligand binding, mitogen-activated protein kinase phosphorylation and de-phosphorylation, and larger models of the epidermal growth factor receptor (EGFR) pathway. Inspired by these computational results, we tested the idea that pulses of EGF could help elucidate the relative contribution of different feedback loops within the EGFR network. These experimental results suggest that models from the literature do not accurately represent the relative strength of the various feedback loops in this pathway. In particular, we observed that the endocytosis and feedback loop was less strong than predicted by models, and that other feedback mechanisms were likely necessary to deactivate ERK after EGF stimulation.

Second, chemical kinetic models contain many unknown parameters, at least some of which must be estimated by fitting to time-course data. We examined this question in the context of a pathway model of EGF and neuronal growth factor (NGF) signaling. Computationally, we generated a palette of experimental perturbation data that included different doses of EGF and NGF as well as single and multiple gene knockdowns and overexpressions. While no single experiment could accurately estimate all of the parameters, we identified a set of five complementary experiments that could. These results suggest that there is reason to be optimistic about the prospects for parameter estimation in even large models.

Third, there is no standard formulation for chemical kinetic models of biological signaling. We propose a general and concise formulation of mass action kinetics based on sparse matrices and Kronecker products. This formulation allows any mass action model and its partial derivatives to be represented by simple matrix equations, which enabled straightforward application of several numerical methods. We show that models that use other rate laws such as Michaelis–

Menten can be converted to our formulation. We demonstrate this by converting a model of *Escherichia coli* central carbon metabolism to use only mass action kinetics. The dynamics of the new model are similar to the original model. However, we argue that because our model is based on fewer approximations it has the potential to be more accurate over a wider range of conditions.

Taken together, the work presented here demonstrates that experimental design methodology can be successfully used to improve the quality of mechanism-based chemical kinetic models.

Thesis Supervisor: Bruce Tidor
Title: Professor of Biological Engineering and Computer Science

Thesis Supervisor: Forest M. White
Title: Associate Professor of Biological Engineering

This thesis is dedicated to
Courtney, Mack, and Boo Boo.
Without your love, support, and sacrifice,
none of this would have been possible.

Acknowledgments

I would like to gratefully acknowledge:

The Members of the Tidor Lab

Bruce Tidor, Jason Biddle, Yuanyuan Cui, David R. Hagen, Bo Kim, Bracken M. King, Gil Kwak, Pradeep A. Ravindranath, Yang Shen, Nate Silver, Kelly M. Thayer, Jared E. Toettcher, David Witmer, Michael Altman, Jay Bardhan, Mark Bathe, David Green, Brian Joughin, and Shaun Lippow

The Members of the White Lab

Forest M. White, Abhinav Arneja, Scott Carlson, Leo Iwai, Emily Johnson, Hui Liu, Emily Miraldi, Jason Neil, Amy M. Nichols, Hideshiro Saito-Benz, Bryan D. Owens, Hideshiro Saito-Benz, Amanda Weaver, Claudia Donnet , Paul Huang, Ji-Eun Kim , Debby Pheasant, Aleksandra Nita-Lazer, Katrin Moser, Jawon Seo, Alejandro M. Wolf-Yadlin, and Yi Zhang

The Members of the MIT Computational and Systems Biology Community

Kristen Naegle, Lauren Frick, and all of the ICBP and CDP members, as well as the Biological Engineering class of 2003.

Specific Acknowledgments

The work on parameter estimation was in collaboration with David Witmer, a very talented UROP. David Hagen has extended these methods to include aggregated rate law models. The work on *E. coli* modeling was in collaboration with Rafel Costa, Yuanyuan Cui, and Stephanie Paige. Kronecker Bio was developed in collaboration Jared E Toettcher, Jaydeep P. Bardhan, and Bo S. Kim.

Funding

This work was partially supported by the National Institutes of Health (P50 GM58762 and U54 CA112967).

Contents

List of Figures

List of Tables

List of Schemes

Chapter 1

Introduction

Despite the fact that biological data is becoming increasing quantitative, more often than not biological knowledge is represented as cartoon drawings or text descriptions. These types of representations are subjective, open to a range of specific interpretations, and difficult to use in a systematic way. In contrast, computational models have the potential to encapsulate our understanding of complex biological processes within a quantitative framework [97, 101]. The promise of these efforts is being able to predict, a priori, the result of an experiment or a clinical intervention. To this end, there has been a strong push to build dynamical models, often based on systems of ordinary differential equations that are capable of recapitulating the kinetic behavior of signaling and metabolic networks. The hope is that these models will lead to a deeper understanding of complex biological processes, facilitate new clinical therapies based on rational design [151, 130], and provide a design substrate for the emerging field of synthetic biology [54]. However, the success or failure of these models is greatly influenced by the data that is used to construct, calibrate and validate them. There is currently a lack of basic methods and best practices for the implementation of models, as well as the design of experiments used in model calibration and discrimination. The goal of this thesis work is to address some of these deficiencies.

1.1 Biochemical Mechanisms Are Only Partially Known

Traditionally, the individual reaction mechanisms that underlie these models have been elucidated individually by genetic, biochemical and biophysical techniques. However, newer high-throughput techniques exist for the large-scale identification of proteins that may be involved in a pathway. Some of these methods directly probe interactions. For example, yeast two-hybrid [61], protein chips [90, 128], and affinity purification [44] look to find pairs of proteins that bind to each other [22]. However, the quality of models based solely on these types of measurements is suspect. In the case of the direct methods, there is a great deal of disagreement between, and often within, data sets [176]. Other approaches are more indirect and find putative interactions by analyzing stimulus response data. Examples of this kind of approach are [65, 4, 161, 147]. In general, these types of approaches suffer from two main shortcomings. First, they rely on statistical methods that need large numbers of replicate experiments to provide accurate answers. These types of replicates are seldom available in biological experiments. One notable exception where this kind of analysis has found a foothold is in the analysis of single cell data as in FACS data [147].

As a result of the limitations of these methods, it is possible if not probable for multiple mechanistically distinct models to fit all of the available data and known chemistry. One way in which mechanisms are tested involves stimulating a system with a step change in the input (adding a high concentration of ligand) and then measuring the change of network outputs (concentration or activities of the various downstream species). If a set of parameters can be found such that the candidate models fits to the data it is taken as a validation of the model. For a linear system this type of experiment provides, at least in a technical sense, enough information to fully identify the system. However, even simple biochemical systems are nonlinear, and as such there is no a priori reason to believe that a step response experiment is sufficient to uncover the relevant dynamics of the system and allow for the selection of a unique model. A simple example of this is the mechanism of antibody–ligand binding studied by Lipschultz, Li, and Smith-Gill [111]. This system consisted of only two parts, a monovalent antibody and its antigen. Even with many time points and accurate measurements it was

impossible to distinguish between these two mechanisms in a standard experiment. However, through the use of a dynamic stimulus Lipschultz *et al.* were able to show that the reaction proceeded according to the two–step mechanism. In this case, the input signal was solved though intuition; however, for more complicated systems a computational approach is needed.

One formulation of this problem is to find the input signal stimulus that maximizes the difference of the outputs of the two candidate models [42]. This approach has been applied successfully to the model selection of a small model (16 equations) of an L-Valine Fermentation process [28]. The fact that these methods have only been applied to relatively small systems highlights the major limitation on these types of methods, which is the computational difficulty of solving the nonlinear dynamic optimization problem. However, in practical experimental design, the global optimum experiment is not needed. A method finds experiments that are good–enough and scales to larger systems would be very useful in practice.

In this work, we address the problem of resolving model ambiguity by providing a method for designing dynamic stimuli that, in stimulus–response experiments, distinguish among parameterized models with different reaction mechanisms. We develop the approach by presenting two formulations of a model-based controller that are used to design the dynamic stimulus. In both formulations, an input signal is designed for each candidate model and parameterization so as to drive the model outputs through a target trajectory. The quality of a model is then assessed by the ability of the corresponding controller, informed by that model, to drive the experimental system. We evaluated our method on models of antibody–ligand binding, mitogen-activated protein kinase (MAPK) phosphorylation and de-phosphorylation, and larger models of the epidermal growth factor receptor (EGFR) pathway. For each of these systems, the controller informed by the correct model is the most successful at designing a stimulus to produce the desired behavior. Using these stimuli we were able to distinguish between models with subtle mechanistic differences or where input and outputs were multiple reactions removed from the model differences. An advantage of this method of model discrimination is that it does not require novel reagents, or altered measurement techniques; the only change to the experiment is the time course of stimulation. Taken together, these results provide a strong basis for using designed input stimuli as a tool for the development of cell signaling models.

1.2 Applying Experimental Design Methodologies to Epidermal Growth Factor Signaling

Motivated by these computational results, we tested the idea that a time varying EGF stimulation could reveal more of the subtle regulation of the EGFR network and help elucidate the relative effects of the different feedback loops. In this work, we took a combined experimental and computational approach. First, we built mechanistic models of the pathway. Second, we used those models to help motivate specific experiments that could be performed in a cell culture based assay. Finally, we used out models, and other models from the literature to help interpret and reconcile the results.

The epidermal growth factor receptor (EGFR) system is arguably the best understood receptor system [164]. It has been extensively studied and modeled by both traditional and high throughput methods[45], and numerous computational models have been constructed to describe various aspects of the signaling [21, 43, 18, 149, 31, 27, 95, 152, 59]. This makes EGFR signaling the perfect substrate for our modeling and computational studies. The emerging picture of this pathway is that it is governed by a complex network of positive and negative feedback loops [185, 45, 3, 146, 116, 124]. Modulation of the strengths of these loops has been implicated in the regulation of pattern formation during development [140, 143], carcinogenesis [82], and cell motility [117, 91].

EGFR (also know as Erb1 and HER) and its family members (Erb2, Erb3, and Erb4) are known to mediate cell-cell interactions in organogenesis and adult tissues [33]. Overexpression of EGFR and Erb2 is a marker of certain types of cancer, including: head, neck, breast [188], bladder and kidney. Overexpression of Erb2 is correlated with poor clinical outcome in ductal breast cancers [184]. As a result of its clinical importance, the EGFR receptors themselves, as well as various down stream proteins, are the targets of therapeutic interventions [76, 158].

Despite clinical interest in the EGFR pathway and over 40 years of intense study, there is still a lot we do not know about the pathway. For example, two recent studies, [19] and [191], found a number of proteins that changed phosphorylation state in response to EGF stimulation that were not previously known to be part of the pathway. Even just at the level of the receptor

there is a large diversity of adaptor proteins that are not understood at a computational level [90]. In addition, many of the known proteins and interactions are not part of any computational model [132].

At the time this work began, the most complete model of EGFR signaling was the model by Hornberg et al. [86]. This model is a refinement of earlier models of the pathway [95, 152, 59]. It describes signal transduction all the way from EGF binding to the receptor to phosphorylation of ERK. The molecular processes include association and dissociation, as well as phosphorylation, dephosphorylation, synthesis, degradation, and trafficking events, all described with mass-action kinetics. Recently, two new models have been published that incorporate additional Erb family members [43, 18], as well as a connection to insulin receptor signaling [21], and neuregulin signaling [149].

However, current models of EGFR signaling, like many cell signaling models, were developed to describe the steady state and step response behavior of the network [43, 95, 152, 59, 21, 86]. Less attention is paid to how the systems recover after the initial stimulus. For example, in the model by Hornberg *et al.* several proteins may be degraded but cannot be synthesized. As a result, the dynamics of these models are dominated by short term signaling events especially the immediate negative feedback loop.

The work presented here will show computationally and experimentally that current models of EGFR signaling miss important dynamics. Moreover, we will show that simple changes to the experimental protocols such as time varying stimulation can help to uncover these missing dynamics. We hope that this will motivate future modeling efforts that will more realistically capture the diversity of feedback mechanisms.

In particular, we tested the mechanistic assumptions of the Hornberg model [86] that lead to the dominance of the endocytic pathway. We stimulated the network with pulses of EGF separated by different lengths of time. The results show that the Hornberg model does not accurately predict the response of the network to this time varying stimulation. We also show that some simple changes previously suggested by us [5] do lead to improved predictions, but ultimately the mechanism previously proposed is not correct. We conclude this work by hypothesizing some alternative mechanisms and suggesting some future experiments that could

help to unravel the details of this complex multi-feedback network.

1.3 Chemical Kinetic Models Contain Many Unknown Parameters

Chemical kinetic models contain many unknown parameters, at least some of which are difficult to measure directly, and instead are estimated by fitting to time-course data. A typical experiment in model calibration involves stimulating a system with a step change in the input (adding a high concentration of ligand) and then measuring the change of network outputs (concentration or activities of the various downstream species). Candidate models are fit to the data and the best model is selected based on the quality of the fit. Most often the calibration problem is formulated as a nonlinear least squares optimization problem, where the goal is to find the set of parameters that minimizes the least squares cost function:

$$\chi^2(\mathbf{p}) = \frac{1}{2n_s n_c} \sum_s \sum_c \frac{1}{T_c} \int_0^{T_c} \frac{[y_{s,c}(\mathbf{p},t) - y_{\text{data } s,c}(t)]^2}{\sigma^2_{s,c}(t)} dt \tag{1.1}$$

where n_c is the number of experimental conditions, n_s is the number of species for which measurements are available, the indices c and s run over the conditions and species, respectively, T_c is the length of the time course for condition c , $y_{s,c}(\mathbf{p},t)$ is the model output for species s and condition c at time t with parameter set$\mathbf{p}$, $y_{\text{data } s,c}(t)$ is the corresponding experimental measurement, and $\sigma^2_{s,c}(t)$ is a weighting factor that is often taken as proportional to the uncertainty of the experimental measurement.

There is a significant amount of work devoted to how best to solve this optimization problem for biological models [10, 89, 125, 144]. In addition, work has been done on optimizing experiments to improve the optimization problem [38, 144] This thesis looks at a related question: Once the parameters are found, how accurately are they determined? One possibility is that there is a single optimum value for parameters as determined by the data. This is the case the model is termed identifiable [47]. Another possibility is that there is a unique optimum, but that many of the parameters around the optimum are "good-enough". If the range of allowable

parameters is large, the model is called practically unidentifiable. It is widely believed that models of systems biology are practically unidentifiable [84, 179, 78, 67, 144].

Previous work by Gutenkunst *et. al* has suggested that even with precise data sets, many parameters are unknowable by measurements of time-course data [78]. What has not been fully explored in these analyses is to what extent experimental design can improve the question of parameter identifiably. One work that examines the effect of time varying stimulation on the parameter optimization is by Faller *et al.* [56]. In that work, they found that the experimental conditions could have a profound effect on the practical parameter identifiably. In Chapter 4 we present a method that is based on genetic manipulation of various proteins. Like Faller, we find that through experimental design we are able to show that a model previously thought to be practically unidentifiable can in fact be identified to high accuracy in only five experiments.

We examined this question in the context of a pathway model of epidermal growth factor (EGF) and neuronal growth factor (NGF) signaling. Computationally, we generated a palette of experimental perturbation data that included different doses of EGF and NGF as well as single and multiple gene knockdowns and overexpressions. While no single experiment could accurately estimate all of the parameters, experimental design methodology identified a set of five complementary experiments that could. These results suggest that there is reason to be optimistic about the prospects for parameter estimation in even large models, that the success of parameter estimation is intimately linked to the experimental perturbations used, and that experimental design methodology is important for parameter fitting of biological models.

1.4 A Standard Formulation for Chemical Kinetic Models

Data formats such as Systems Biology Markup Language (SBML) [62, 88] and Cell Markup Language (CellML) [112] have standardized the representation of biological models, allowing them to be shared between laboratories and encouraging the development of model repositories. This has greatly increased the number of publicly available models [129]. However, unlike the models themselves, computational methods have remained difficult to share. For the most part they rely on one–off, and often proprietary, codes. As a result the methods have remained siloed

within individual software packages and are difficult to combine, customize, or apply in new contexts. The problem is exacerbated by the fact that CellML and SBML allow for arbitrary mathematical expressions for the rate laws. Designing general codes for the analysis of models that may contain arbitrary mathematical structure leads to high computational overhead, or spotty support.

The difficulty in implementing reusable computational tools arises, in part, from the fact that there is no standard formulation for chemical kinetic models of biological signaling. This makes it difficult implement general-purpose analytical and computational methods. Here, we propose a general and concise formulation of mass action kinetics based on sparse matrices and Kronecker products. This formulation has the advantage that the matrices are constant and depend linearly on the rate parameters, and that the quadratic terms are expressed as Kronecker products. We show that this formulation allows simple algebraic manipulation of the models and can serve as the basis for efficient general-purpose codes. In particular, we formulate numerical integration, steady state finding, sensitivity analysis, and model calibration. In this work, we applied these methods to a series of models which demonstrated that they scale to models of thousands of equations on standard desktop hardware, despite being implemented in a high-level mathematics package [168].

1.4.1 The Law of Mass Action

Our formulation for chemical kinetics is based on the Law of Mass Action. The observation that the rate of a chemical reaction is related to the concentration of the reactants dates back to Wenzel in 1777 [105]. In 1867 Guldberg and Waage put this observation into a quantitative form, noting that the rate of a reaction was proportional to the products of the concentrations of the reactants [178]. Subsequently, it was showed by van't Hoff that this law could be derived from the principles of thermodynamics [118].

Zeroth-order reaction:

$$\emptyset \xrightarrow{r} A \qquad \frac{dA}{dt} = r \tag{1.2}$$

First-order reaction:

$$A \xrightarrow{r} B \qquad -\frac{dA}{dt} = \frac{dB}{dt} = r\ A \tag{1.3}$$

Second-order reaction:

$$A + B \xrightarrow{r} C \qquad -\frac{dA}{dt} = -\frac{dB}{dt} = \frac{dC}{dt} = r\ A\ B \tag{1.4}$$

A modern interpretation of chemical kinetics can be derived from statistical mechanics [120, 71, 72]. The basic assumption is that the probability of a reaction occurring is determined by the concentration of reacting molecules. Put more precisely, given a system with n copies of a molecule X there is some probability that over a fixed time step δt a reaction will occur to produce one more or one fewer molecules of X. We call the probability of forming one more molecule f_n and one fewer r_n. Likewise, for systems with $n-1$ or $n+1$ reactants there are forward and reverse probabilities f_{n-1}, r_{n-1} and f_{n+1}, g_{n+1} respectively. If we let p_n be the probability that there are n molecules of X then the time evolution of p_n is given by Equation 1.5.

$$\frac{dp_n}{dt} = -(f_n + g_n)p_n + (f_{n+1} + g_{n-1})p_{n-1} + g_{n+1}p_{n+1} \tag{1.5}$$

If we are interested in the average number of molecules, then Equation 1.5 becomes:

$$\frac{d}{dt}\langle n \rangle = \langle \mathrm{f_n} \rangle - \langle g_\mathrm{n} \rangle\,. \tag{1.6}$$

Returning to our assumption that the rates of reactions are given by the products of the concentrations of their reactants we have:

$$\begin{array}{lll} \emptyset \rightarrow \text{products} & f = \frac{k}{v} & \langle f_n \rangle = k\frac{1}{v} \\ \mathrm{A} \rightarrow \text{products} & f_n = \frac{k}{v}n & \langle f_n \rangle = k\frac{\langle n \rangle}{v} \\ \mathrm{A} + \mathrm{B} \rightarrow \text{products} & f_{n,m} = \frac{k}{v}nm & \langle f_n \rangle = k\frac{\langle nm \rangle}{v} \approx k\frac{\langle n \rangle}{v}\frac{\langle m \rangle}{v} \end{array} \tag{1.7}$$

Often this average is referred to as the deterministic limit of the chemical master equation [71]. From this, we can see that our original empirical description of the kinetics represents a kind

of average limit of a Poisson process. Even though many biochemical species are present in low numbers within a cell, variation about this average due to small number statistics is typically not a major factor as biological data is averaged over many cells, each of which can be though of as an independent realization of this Poisson process. However, a subtle approximation made in Equation 1.7, that n and m are uncorrelated, keeps this relationship from being exact. In cases of bifurcations, then, the deterministic solution may not describe the average even in the limit of large numbers of molecules or realizations [71]. For the models considered here there is no multistability so we opted for the deterministic solution. However, an interesting extension to this work would be to apply it to systems where the stochastic solution is necessary.

1.4.2 The Quasi-Steady State Approximation

The style of modeling described above, which is based on networks of elementary reactions, is not the most popular style of modeling. In fact, the majority of models in the literature use additional levels of approximation. For example, the stepwise reaction [119] by which an enzyme binds a substrate, modifies that substrate, and then releases it is represented as a single aggregated-reaction with a Michaelis–Menten rate law [123, 26]. We call these models aggregated rate law (ARL) models to distinguish them from the mass action rate law (MRL) models described above.

One class of simplification is based on separation of time scale arguments such as rapid equilibrium [154, page 18], or quasi-steady state [154, page 505]. Another class is based on approximations that hold over certain concentration regimes such as pseudo–first-order kinetics. More complex multi-step or multi-enzyme mechanisms may be aggregated into a small number of pseudo-reactions [74, 11, 69, 110]. Often these represent empirical fits rather than a physical limit of a specific mechanism [192]. While aggregated rate laws (ARLs) generally simplify the set of reactions, they increase the mathematical complexity of the rate laws which, in general, are non-linear functions of substrate, product, activator, and inhibitor concentrations. Moreover, ARL based models are valid only over a particular concentration and time regime [172, 154].

An example of an ARL is the Michaelis–Menten rate law for a uni–uni irreversible enzyme.

The chemical mechanism is shown in Equation 1.8:

$$\mathrm{E} + \mathrm{S} \underset{k_{-1}}{\overset{k_{+1}}{\rightleftharpoons}} \mathrm{E{:}S} \xrightarrow{k_2} \mathrm{E} + \mathrm{P} \tag{1.8}$$

The Michaelis–Menten rate law describes the rate of formation of product as a function only of the substrate concentration. The range of validity of this rate law is derived in detail in [155]. One requirement is that approach to equilibration for the complex is fast compared to the production of product:

$$\frac{E_{\text{active}}(t)}{K_{\text{M}} + S_{\text{total}}(t)} \ll \left(1 + \frac{k_{-1}}{k_2}\right)\left(1 + \frac{S_0}{K_{\text{M}}}\right) \tag{1.9}$$

However, in the case of a signaling cascade, the total amount of substrate ($S_{\text{total}}(t)$) and the total amount of active enzyme ($E_{\text{active}}(t)$) change dynamically, and in many cases quite rapidly. Therefore, for the approximation to hold the approach to equilibrium must also be fast compared to the modulation of the enzyme and substrate by other mechanisms in the model.

The second criteria is that the amount of free substrate is approximately equal to the total amount of substrate $S(t) \approx S_{\text{total}}(t)$. Put another way, it is assumed that the enzyme–substrate complex does not significantly deplete the free substrate.

$$1 \gg \frac{E_{\text{active}}(t)}{K_{\text{M}} + S_{\text{total}}(t)} \tag{1.10}$$

However, in networks where the enzymes and their substrates are both proteins it is very common for the concentration of the enzyme to exceed the concentration of the substrate and its K_{M}. For example, Ras is a kinase, which phosphorylates Raf. In HeLa cells, the concentration of Ras is approximately 0.40 μM and the concentration of Raf is 0.013 μM [66]. Estimates of the K_{M} range from 0.0053 to 0.1 μM. Even at the highest value of K_{M} condition 1.10 doesn't hold. Moreover, if this condition holds for a particular cell type, overexpression or knockdown studies could push the system out of the regime where the approximation holds. It is for these reasons that we prefer to use the elementary reaction formulation as it applies over all concentration and time scales.

It is worth noting that while the models developed in this thesis are elementary mass action rate laws, the experimental design methods developed in Chapters 2 and 4 apply to models of both types.

1.4.3 Converting Non-Mass Action Models

One practical challenge in building mass action rate law (MRL) based models is the lack of data available in the literature to parameterize these models. Much of the parameter data that is available is expressed in terms of the parameters of aggregated rate laws (ARL)s, such as $v_{\max}$, K_{M}, and K_{I}. Here we present a method that can automatically derive the elementary mass action rate constants from these ARLs. This method is a significant improvement over previous methods. It is able to accommodate a wider range of ARLs, and is computationally more efficient. We apply these methods to construct a MRL of the *Escherichia coli* central carbon metabolism based on an ARL model from the literature [41]. Results are presented that show the two models perform similarly over the range of inputs for which the ARL model was developed. Moreover, we argue that because the MRL model is based on fewer approximations it has the potential to be accurate over a wider range of conditions.

1.5 Anticipated Impact of this Work

Generally speaking, development of these models can be divided into three tasks. First, the model variables have to be selected. Usually this is the set of chemical species involved in the system. Second, the model topology is determined. This is the set of physical and chemical interactions that create, destroy, move, and interconvert the chemical species. Third, the models usually contain many unknown parameters which must be set. Typically, this involves fitting kinetic rate parameters and initial concentrations to experimental data [173]. This process is often called parameter fitting or model calibration. In addition to these three steps, there is an implicit zeroth-order decision, which is selecting the level of detail to model. This choice will dictate how the various interactions will be represented mathematically.

The standard approach to model building is shown schematically in Figure 1-1. This method

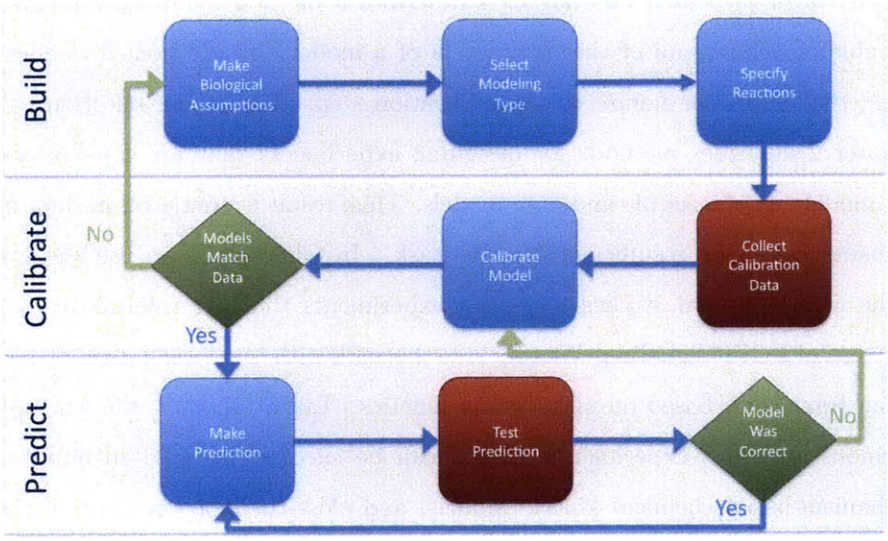


Figure 1-1: Schematic of the Standard Model Building Process

begins with a description of the biological process to be modeled. The model is then calibrated to all of the available data. In some cases additional experiments may be performed to supplement this data. Once a parameter set is found that can describe all of the available data, it is used to make predictions about the experimental system. The model is considered "validated" if the predictions agree with experimental observation. Implicit in this diagram is that if a model matches all of the available data, it is allowed to pass to the next step of the process. For example, if a set of parameters can be found such that the model matches the data available, it is said to be calibrated and it is passed onto the prediction step. There is no notion of there being sufficient data to fully calibrate the model, or that any given step is carried to completion.

The methods and results presented here are an attempt to close-the-loop on the various

steps of this process. Our modified model building process is shown in Figure 1-2. The model calibration work presented in Chapter 4 describes a method for designing experiments that are capable of defining all of the parameters of a model. In addition, it defines a metric for assessing the success or failure of the calibration step. The model selection work presented in Chapter 2 describes methods for designing experiments that are capable of selecting the correct model out of a set of candidate models. This, result is nontrivial as data from standard experiments would be insufficient for this task. In Chapter 3, we use the insights gained from the computational studies to design experiments that are informative of the detailed mechanisms of EGFR signaling. We conclude the work with the development of a computational modeling framework based on mass action kinetics. Taken together, the work presented here will demonstrate that experimental design can be successfully used to improve the quality of mechanism-based chemical kinetic models, and close-the-loop on some of these modeling decisions.

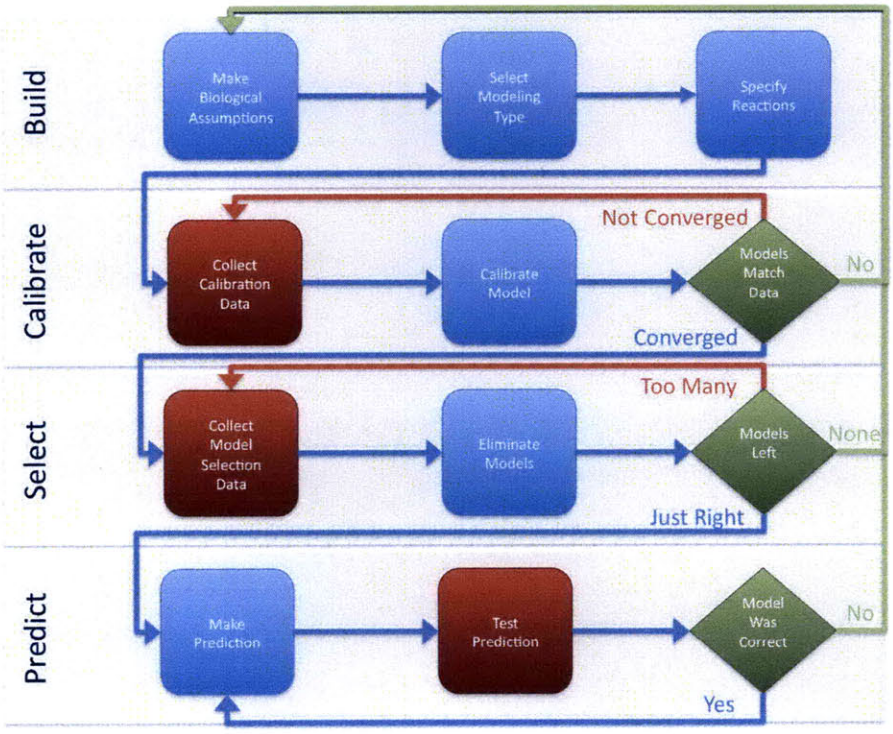


Figure 1-2: The Closed Loop Model Building Process. The work presented here provides methods for designing the experiments at each step, as well as conditions for determining if a step has been completed.

Chapter 2

Stimulus Design for Model Selection and Validation in Cell Signaling

2.1 Introduction

One goal of systems biology is to develop detailed models of complex biological systems that quantitatively capture known mechanisms and behaviors, and also make useful predictions. Such models serve as a basis for understanding, for the design of experiments, and for the development of clinical intervention. In support of this goal, there has been a strong push to build mechanistically correct kinetic models, often based on systems of ordinary differential equations (ODEs), that are capable of recapitulating the dynamic behavior of a signaling network. These models hold the promise of connecting biological and medical research to a class of computational analysis and design tools that could revolutionize how we understand biological processes and develop clinical therapies [97, 101].

One type of experiment for model validation involves stimulating a system with a step change in the input (typically by adding a high concentration of ligand) and then measuring the change of network readouts (the concentrations or activities of the various downstream species) as a function of time. Candidate models are fit to the data and the best model is selected based on criteria such as the quality of the fit, the simplicity of the model, and other factors. While it is

tempting to select a simple model consistent with the known biochemical mechanisms that fits all available data, future experimentation may prove this choice incorrect. Rather, it may be preferable to collect "all" models consistent with known mechanisms and data, and to design follow-on experiments capable of distinguishing among the model candidates. Here we develop an approach for designing these experiments using dynamic stimuli.

While the step-response experiment is attractive for its ease of implementation, dynamic stimuli have the potential to uncover more subtle system dynamics and to improve model selection in the cases where step-response experiments are not sufficiently discriminating. One example that illustrates the use of a dynamic stimulus to distinguish between two models is the work by Smith-Gill and co-workers on the detailed mechanism of antibody-antigen binding [111]. Initial step-response experiments were compatible with either a one-step or two-step binding mechanism, in which the ligand and antibody first come together in a loose encounter complex before forming a fully bound complex. To resolve this ambiguity, the authors applied a series of rectangular pulses of ligand concentration to their system. The resulting binding curves produced by this dynamic stimulus were inconsistent with the one-step model but were consistent with a two-step model and suggested the existence of an encounter complex, even though such a complex could not be measured directly by the assay.

These results show that time varying inputs have the potential to distinguish closely related models of biochemical systems. For the relatively simple antibody-antigen system, an appropriate dynamic input was deduced intuitively. However this sort of intuitive design is difficult, especially in the case of more complex cell signaling pathway models, which may be described by hundreds or thousands of differential equations. An automated approach that could design experiments to test these complex systems has the potential to expand the scope of model selection experiments.

Previous work in designing dynamic stimuli for the purpose of model discrimination in systems biology has focused on choosing input trajectories that maximize the expected difference in the output trajectories of competing models [23, 8, 48, 127, 29, 42, 100]. In addition to model discrimination, a rich literature exists on experimental design in systems biology for the purpose of estimating model parameters [101, 89, 174, 38]. These optimization approaches for model

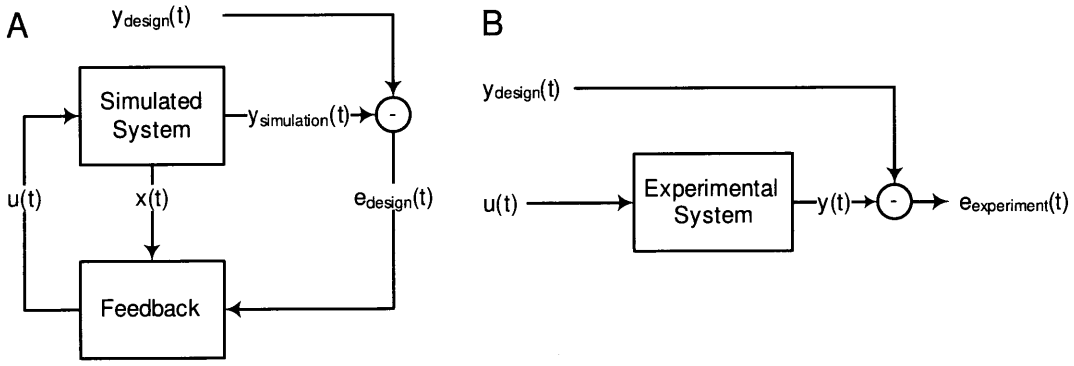


Figure 2-1: Schematic of Experiment Design. (A) A feedback controller is used to solve for the stimulus $u(t)$ that will drive the model system outputs $y_{\text{simulation}}(t)$ to follow the design trajectory $y_{\text{design}}(t)$. The inputs to the feedback controller are the deviation from the desired trajectory $e_{\text{design}}(t)$ as well as the model state $x(t)$. (B) The designed stimulus can be applied to an unknown experimental system to assess the quality of the model. A stimulus based on a good model should be able to drive the experimental system output $y(t)$ through the design trajectory.

discrimination have been applied to small biological systems, but the nonlinearity of the models combined with the presence of many local minima has thus far limited their application [29].

There is a need to extend these methods to design experiments that may not be optimal but are capable of discriminating between large pathway models. Instead of trying to design an input signal that maximizes the predicted difference between two model readouts, we recast the problem as a control problem (Figure 1). We choose a target trajectory, and then challenge a model-based controller to drive the system to follow the target trajectory. The extent to which the controller based upon a given model is able to drive the physical system is a measure of the fitness of that model.

We demonstrate our methodology by applying it to the epidermal growth factor receptor (EGFR) pathway. This pathway has been extensively studied and modeled [25, 181, 86, 95, 152]. EGFR and its family members (Erb2, Erb3, and Erb4) are known to mediate cell-cell interactions in organogenesis and adult tissues [33]. Overexpression of EGFR family members is a marker of certain types of cancer, including head, neck, breast, bladder, and kidney [184]. Because of their clinical importance, the EGFR receptors themselves, as well as various down-

stream proteins, are targets of therapeutic intervention [122, 53]. Despite clinical interest in the EGFR pathway and over 40 years of intense study, there is still much about the pathway that is not known. For example, in three recent studies [191, 19, 133], a number of proteins that changed phosphorylation state in response to EGF stimulation were found that were not previously known to be part of the pathway. In addition, many of the known pathway proteins are not part of any computational model [132].

The ordinary differential equation model of Hornberg et al. is a widely used mechanistic model of EGFR signaling [86]. This model is a refinement of earlier models of the pathway [95, 152, 59]. It describes signal transduction initiated at the surface by EGF binding to EGFR, leading eventually to the dual phosphorylation of ERK as the most downstream outcome. ERK then participates in a negative feedback to the top of the pathway. The elementary molecular processes modeled include bimolecular association and dissociation, phosphorylation and dephosphorylation, synthesis and degradation, as well as endocytosis and trafficking all described with mass-action kinetics. The model contains 103 chemical species, 148 reactions, 97 independent reaction rates, and 103 initial conditions.

We applied our computational methods initially to a small portion of the EGFR model for development and demonstration purposes, and then to the full model. In both cases, we formulated a set of closely related models that exhibit similar step-response behavior. We built a controller capable of controlling each candidate model and asked the controller to drive the system output (doubly phosphorylated ERK) to a predetermined value. Finally, by applying these designed inputs based on the reference and perturbed models, we show that it is possible to discriminate between the various model alternatives.

2.2 Methods

2.2.1 Model Formulation

In this work, we consider mass-action kinetic models consisting of zeroth-, first-, and second-order reactions described by ordinary differential equations. In the equations below, k signifies a rate constant; A , B, and C represent species or concentrations of species, depending on the

context; and $\emptyset$ is the empty set or nothing.

Zeroth-order reaction:

$$\emptyset \rightarrow A \qquad \frac{dA}{dt} = k \tag{2.1}$$

First-order reaction:

$$A \rightarrow B \qquad -\frac{dA}{dt} = \frac{dB}{dt} = kA \tag{2.2}$$

Second-order reaction:

$$A + B \rightarrow C \qquad -\frac{dA}{dt} = -\frac{dB}{dt} = \frac{dC}{dt} = kAB \tag{2.3}$$

Large systems of reactions of this form can be represented compactly using Equation 2.4.

$$\begin{aligned} \frac{dx}{dt} &= A_1 x + A_2 x \otimes x + B_1 u + B_2 x \otimes u + k \\ y &= Cx \end{aligned} \tag{2.4}$$

The state vector x describes the chemical species concentrations that are free to evolve in time according to the kinetics of the system. The input vector u represents the chemical species concentrations controlled by the experimenter. Matrices A_1 and B_1 represent first-order reactions, matricesA_2 and B_2 represent second-order reactions, and k represents constitutive (zeroth-order) reactions. The symbol $\otimes$ denotes the Kronecker Product (also known as the matrix direct product) [75]. For vectors, this operator generates a vector of all quadratic products.

$$x \otimes u = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \otimes \begin{bmatrix} u_1 \\ \vdots \\ u_m \end{bmatrix} = \begin{bmatrix} x_1 u_1 \\ \vdots \\ x_1 u_m \\ x_2 u_1 \\ \vdots \\ x_n u_m \end{bmatrix} \tag{2.5}$$

The output of the model y is a linear combination of the state variables represented by the matrix$C$.

2.2.2 Control Formulation

A controller was developed to solve for the input signal $u(t)$ that best achieves a particular objective in the output. We formulate this objective as a cost function $G(u)$ that measures the distance between the model output and the desired output.

$$G(u) = \int_0^T [y(u,t) - y_{\text{design}}(t)]^2 \, dt \tag{2.6}$$

Here, $G(u)$ is the sum of squares error between $y(u,t)$, the model output for a given input $u(t)$, and $y_{\text{design}}(t)$ the target output the controller is trying to match. T is the length of the experiment. The control problem is then to find an input function $u(t)$ that minimizes $G(u)$.

Equation 2.6 depends on models of the form of Equation 2.4, which are nonlinear and potentially high-order. This prevents us from solving the minimization problem directly. To address this issue, we implement two different approximations. The first is based on controlling a model formed from successive linearizations of Equation 2.4 (henceforth referred to as the tangent linear controller), and the second is based on a local search of the input space (henceforth referred to as the dynamic optimization controller) [36].

2.2.3 Tangent Linear Controller

A first-order approximation to Equation 2.4 at time t was computed by taking the Taylor series expansion about the current value of the state and input vectors (x_t and u_t).

$$\begin{aligned} \frac{d}{dt}\Delta x &\approx \left[A_1 + A_2(I \otimes x_t + x_t \otimes I) + B_2 I \otimes u_t\right]\Delta x + \left[B_2 x_t \otimes I\right]\Delta u \\ y &\approx C\left(x_t + \Delta x\right) \end{aligned} \tag{2.7}$$

Equation 2.7 is a linear differential equation with state variable Δx and time varying forcing term Δu, which has both numerical and analytical solutions. However, this approximation would tend to diverge from the solution to Equation 2.4 as Δt, the time beyond the linearization point t, and as$(\Delta x, \Delta u)$, the distance from the linearization point (x_t, u_t), increases. To mitigate this problem the true system (Equation 2.4) was propagated, and successive linearizations were applied to improve the controller performance. Effectively, the linearization point is allowed to slide along with the exact simulation.

Operationally, each time step was solved in three stages. First, the current state of the nonlinear simulation was used to derive a linear approximation about the current time point. Second, the linear system was solved to get the best input Δu. The linear system was solved numerically by discretizing the input as a series of scaled and shifted boxcar functions [177] of width τ. Numerical integration with the MATLAB routine ode15s [156] was used to compute the system response to a unit boxcar input. The output of a linear time invariant system can be expressed as a linear combination of scaled shifted impulse response functions. Thus, solving for the input was achieved by computing the weights to apply to the input pulses that gave the optimal output. This was solved as a linear system of equations with box constraints on the input to limit the maximum and minimum concentration using the MATLAB routine lsqlin. Third, the computed input signal was applied to the full nonlinear system for a short time step τ. The process was the repeated for the next time interval. Effectively, each step the algorithm solves for an input signal Δu that is piecewise constant. The width of the intervals τ as well as the number of intervals is a parameter of the optimization and should be chosen based on the accuracy of the linear system.

2.2.4 Dynamic Optimization Controller

In this controller formulation, rather than exactly solving the tangent linear system, we solved the full nonlinear problem iteratively using a gradient optimization method. Application of this method requires computation of the sensitivities of the least squares objective function (Equation 2.6) with respect to the input parameterization p. An efficient way to compute this quantity is to first solve for the adjoint sensitivities λ [36]. For the dynamical system (Equation 2.4) and the objective function (Equation 2.6), the adjoint equations are given by Equation 2.8.

$$\begin{aligned} \frac{d\lambda^*}{dt} &= \lambda^* \left[A_1 + A_2(I \otimes x + x \otimes I) + B_2 I \otimes u\right] + 2C^T \left(Cx - y_{\text{design}}\right) \\ \nabla_p G(p) &= \int_0^T \lambda^* \left[B_1 + B_2 \left(x \otimes I\right)\right] \nabla_p u(p) dt \end{aligned} \tag{2.8}$$

Here, λ^* indicates the conjugate transpose. For piecewise linear input signals the i^{th} component of the gradient $\frac{du}{dp_i}$ is given by:

$$\frac{du(t)}{dp_i} = \begin{cases} 0 & t \leq T_{i-1} \text{or} t \geq T_{i+1} \\ \frac{(t-T_{i-1})}{T_i - T_{i-1}} & T_{i-1} < t < T_i \\ \frac{(T_{i+1}-t)}{T_{i+1}-T_i} & T_i < t < T_{i+1} \end{cases} \tag{2.9}$$

The adjoint equations were solved in MATLAB using ode15s [156] and the optimization was implemented using fmincon configured to use Quasi-Newton [165] with BFGS [131, 14] in the MATLAB Optimization Toolbox Version 3.1.1.

2.2.5 Constraining Input Signals

Thus far the input signals have been unconstrained, except by the choice of the discretization. However, in practice it may be desirable to restrict the space of input signals to those that could be feasibly achieved by a given experimental setup. For example, in many experimental setups it is easy to add material but difficult to take material away. Likewise, there may be a maximum and minimum concentration for the input signals, or a maximum rate of change for the input

signal. We implemented these experimental constraints as linear inequality constraints of the form of Equation 2.10.

$$Ap \leq b \tag{2.10}$$

The matrix A and the vector b are passed as arguments to lsqlin in the case of the tangent linear controller, or to fmincon in the case of the dynamic optimization controller. An example of a linear constraint that might be applied is that the input increase monotonically. In this case, A and b are given by Equation 2.11.

$$A = \begin{bmatrix} 1 & -1 & & 0 \\ & \ddots & \ddots & \\ 0 & & 1 & -1 \end{bmatrix} \quad \text{and} \quad b = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \tag{2.11}$$

2.2.6 EGFR Signaling Model

We based our model of EGFR signaling on that of Hornberg et al. [86], which itself is a refinement of earlier work [95, 152, 59]. The model contains 103 chemical species, and 148 elementary reactions; these reactions are of the type given by Equations 2.1, 2.2, and 2.3 and may be reversible. The model is parameterized by 97 distinct reaction rate values and 103 initial conditions. The details of this model are given in Appendix A.2.

Here we also introduced a modified model of EGFR signaling, which contained six additional production/degradation reactions of the form of Equation 2.12, where X is one of of the following: GAP, GRB2, SOS, RAS-GDP, SHC, or GRB2-SOS

$$\emptyset \underset{k_{\text{deg}}}{\overset{k_{\text{synth}}}{\rightleftharpoons}} X \tag{2.12}$$

The degradation rate k_{deg} was set such that the steady-state value of the species was the same as the steady-state value in the unmodified model computed using Equation 2.13.

$$k_{\text{deg}} = k_{\text{synt}}/X_{\text{ss}} \tag{2.13}$$

In addition to the protein synthesis and degradation reactions, a GAP-catalyzed turnover of RAS-GTP was implemented.

$$\mathrm{RAS-GTP + GAP} \underset{k_{\mathrm{off}}}{\overset{k_{\mathrm{on}}}{\rightleftharpoons}} \mathrm{RAS-GTP:GAP} \overset{k_{\mathrm{cat}}}{\rightarrow} \mathrm{RAS-GDP + GAP + P_i} \tag{2.14}$$

The rate constants (k_{on}, k_{off}, and k_{cat}) are $5 \times 10^{-7}\mathrm{molecules}^{-1}\mathrm{cell}^{-1}\mathrm{s}^{-1}$, $0.4\mathrm{s}^{-1}$, and $0.023\mathrm{s}^{-1}$, respectively. The rate constants k_{on} and k_{off} are taken from the analogous reaction where GAP is part of the receptor complex and the k_{cat} was fit so that the half-life of RAS-GTP in the absence of EGF matched literature values [113].

Finally, a first-order turnover of internalized SOS was implemented with a rate constant of $10^{-7}\mathrm{s}^{-1}$ based on the turnover rate of EGFR.

$$SOS_i \rightleftharpoons SOS \tag{2.15}$$

This augmented model had the additional property that if the input is removed (set to zero) it will return to its initial condition.

2.2.7 MAPK Signaling Model

The mitogen activated protein kinase cascade is a signaling motif found repeated throughout biology [59]. In each step of the cascade a substrate is multiply phosphorylated by a kinase, which in turn is the input to the next layer in the cascade. The off signal, present in each layer, is a phosphatase that removes the phosphate groups. Despite knowing all of the species involved, the detailed mechanism of the enzymatic steps had been difficult to determine [59]. In particular, it was unclear if the kinase acted in two distinct enzymatic steps, whereby it released the substrate between phosphorylation steps (distributive mechanism) or if it performed both phosphorylation steps before releasing the substrate (processive mechanism).

A MAP kinase cascade consisting or RAF, MEK, and ERK is contained in the Hornberg EGFR pathway model. We extracted a tier of this cascade consisting of a single kinase,

phosphatase, and substrate. The four reversible bimolecular reactions representing the phosphorylation of ERK by doubly phosphorylated MEK (MEKpp) and the dephosphorylation by a phosphatase were used as the basis of a new model. The model contains a distributive dual phosphorylation step catalyzed by MEKpp and a distributive dual dephosphorylation step catalyzed by a phosphatase. MEKpp is the system input; doubly phosphorylated ERK (ERKpp) is the output.

In addition to this basic model, three alternative models were also constructed that differed in their mechanism of phosphorylation and de-phosphorylation (processive or distributive). The set of four models (distributive-kinase – distributive-phosphatase, processive-kinase – distributive-phosphatase, distributive-kinase – processive-phosphatase, processive-kinase – processive-phosphatase) represents all possible combinations of processive and distributive phosphorylation and de-phosphorylation mechanisms (Figure 2-3A). The alternative models, which contain some rate parameters not included in the distributive-kinase – distributive-phosphatase base model, were parameterized by fitting the parameters to the step response of the double distributive model, which included both a step-up and a step-down experiment. The details of these four models are given in Appendix A.1.

2.3 Results

We have developed a method for designing input signals that are capable of controlling the output of a candidate model. In practice, these input signals are useful for distinguishing among sets of candidate models.

2.3.1 Simple Antibody Binding Models

The dynamic optimization controller was applied to design input stimuli for each of the two alternative antibody binding reactions studied by Smith-Gill and co-workers [111]. For both the one-step and the two-step model (Appendix A.1), the objective applied was to produce a constant output of antibody-ligand complex from time zero onwards. In the experiment performed by Smith-Gill and co-workers the measurement was a change in mass due to ligand

binding as measured by surface plasmon resonance. While the fully bound complex is more stable than the postulated encounter complex, both have the same mass and would produce the same output signal. Therefore, in the case of the two-step model, the output is the sum of the encounter and fully bound complexes, whereas in the one-step model it is simply the fully bound complex. The basis set for the input was a 50 point piecewise-linear function with linear spacing. In the two-step model points were distributed evenly over the entire interval. In the one-step model points were placed evenly from 500 s to 600 s to accommodate the sharp transition.

The results are shown in Figure 2-2. Both controllers designed an input signal that starts at high concentration to form the complex quickly and then drops to a lower concentration to keep the complex from overshooting the desired value. However, the controller for the one-step model drops abruptly while the controller for the two-step model drops more gradually. The desired output is not recovered when the stimuli from the wrong models are applied. When the one-step input was applied to the two-step system, the output produced an undershoot followed by an overshoot. When the input designed for the two-step model was applied to the one-step system, the complex concentration also produced an overshoot, but one that persisted. In both cases, accounting for the presence or lack of the encounter complex was critical for controlling the output correctly.

It is interesting to note that this method allows for the selection of both the more complex model (if it is correct) as well as the simpler model. This is not possible using standard *a posteriori* metrics, such as least squares, which will always favor the more complex model. While there are methods that try to correct for this bias [24], properly accounting for model complexity in large nonlinear systems remains an open problem [81]. Comparing our results to the Smith-Gill pulse method (Figure 2-2B), it is clear that both computational experiments permit the two models to be distinguished in favor of the two-step method. However, for larger and more complex cases, it is unclear whether intuitive approaches or square pulse inputs will be sufficient to design distinguishing experiments. Another feature of the simulations is that the designed pulse produces a level output that does not require fine time resolution to accurately measure. This can be a significant advantage for more complex experimental systems, such

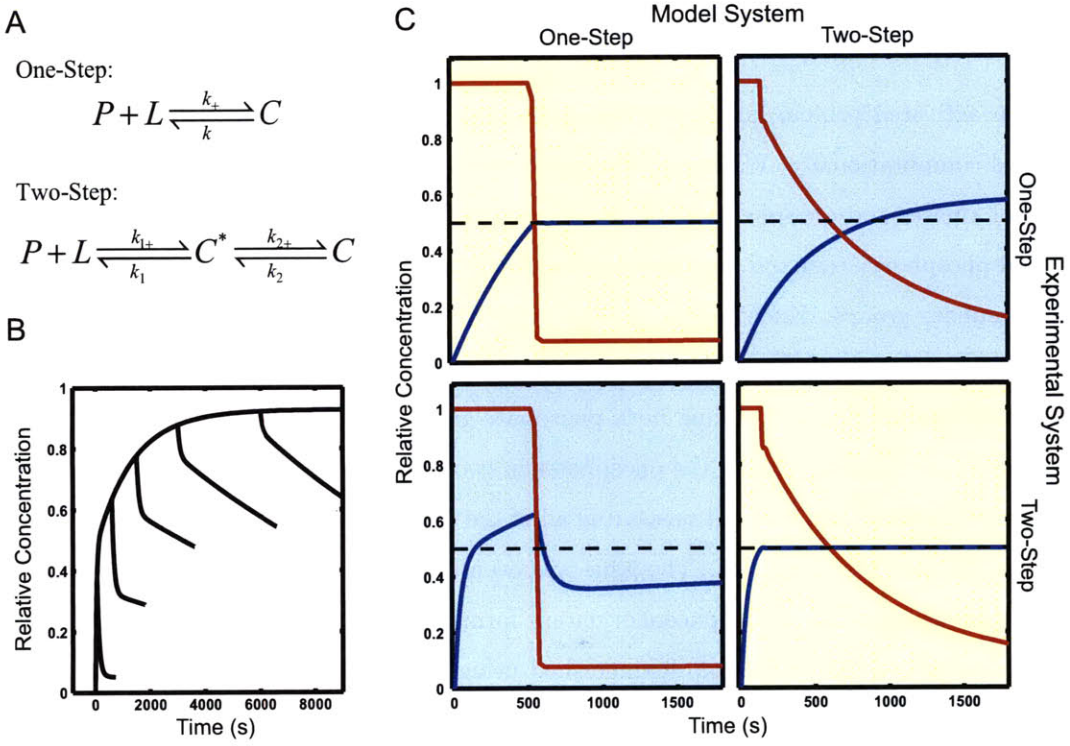


Figure 2-2: Analysis of Monovalent Antibody Binding (A) Two models of monovalent antibody binding, a one-step version with no intermediate, and a two-step version with an association intermediateC^*. (B) The results of six simulated experiments are shown as designed in ref-3. Each trace is the response of the system to a square pulse of ligand concentration. The width of the pulse varies from $400s$ to $6,000s$. The pronounced elbow in the middle curves is indicative of the two-state model. The one-step model cannot have compound off kinetics. (C) The set of experiments designed by this algorithm as well as simulated results are shown. Each pulse was designed to produce a level output when applied to the correct model (yellow boxes), which was observed, and produced a distinctly different result when applied to the other models (blue boxes). The red lines are the inputs (unbound L) the blue lines are the output (C or $C+C*$). The smaller the gap between the blue and the black dashed line the better the model fits the real system. Looking across one row shows a pair of experiments that would be run together.

as cell signaling measurements, where limitations on experimental observations are even more severe, whether in terms of numbers of species, time points or other factors.

2.3.2 MAPK Signaling

Mitogen activated protein kinase cascades have been extensively studied experimentally and modeled computationally. While many variants exist, the canonical pathway consists of three layers of kinases and phosphatases. For each layer, the kinase activates the downstream kinase by dual-phosphorylation and the phosphatase deactivates the downstream kinase by removing the phosphate groups. Knowing the general structure of this pathway, it was still difficult to determine the details of the enzymatic steps. In particular, it was unknown if the kinase acted in a processive mechanism (adding both phosphate groups in a single step), or if it acted in a distributive mechanism (adding the phosphates in two distinct enzymatic steps). The difficulty arose from the fact that, without measuring all of the phosphorylation forms, both mechanisms could fit the step response data. The issue was eventually resolved by devising an experiment that could separate all of the phosphorylation forms [59]. Here we show that, in principle, the mechanisms could have been distinguished using our method, without adding additional measurements.

To address this problem we generated four candidate models of a mitogen activated protein kinase (MAPK) dual phosphorylation reaction. All four models contained forward phosphorylation and reverse dephosphorylation steps, but differed in the detailed mechanisms. For both the forward and the reverse reactions we considered a processive (one-step) and a distributive (two-step) mechanism (Figure 2-3A). Taking all combinations of distributive and processive reactions produced four models. For each model the free kinase concentration was the input variable and the concentration of doubly phosphorylated substrate was the output.

For each of the four models, a stimulus was developed using the tangent linear controller. The objective was to drive the output to a fixed value that remained constant with time. Each of the four designed signals was used to stimulate each of the four models, and the resulting 16 experiments are shown in Figure 2-3C. Along the diagonal, one can see that the input signal

derived from the correct model was able to effectively control the system. However, looking at each off-diagonal entry shows that inputs from each wrong model did a poor job controlling each system. In any real experiment, there is only one true system, which corresponds to performing the experiments from a single row of the figure.

As with the antibody models, the algorithm was able to find a set of signals that distinguished amongst multiple models. It is worth noting that these solutions were generated automatically from the candidate models and did not require explicit user supervision.

2.3.3 EGFR Pathway

A popular ordinary differential equation model of the EGFR pathway is that of Hornberg and co-workers [86]. This model consists of 103 differential equations and includes ligand binding, receptor dimerization and activation, adaptor protein binding, trafficking of the receptor complex, and activation of the MAPK cascade terminating with ERK dual phosphorylation (Figure 2-4). This model was built as a set of successive refinements of earlier models [152, 95, 59], with each refinement adding a new level of detail to the model. In its most recent formulation an additional negative feedback loop was added whereby activated ERK phosphorylates SOS and deactivates it. This model has been shown to agree with time course data collected in cell based assays as well as literature values for parameters measured in vitro [152]. We compare the original Hornberg model to a version with additional changes. We continue this model evolution by modifying the Hornberg model so that, when the input (EGF) is removed, the model returns to its initial conditions. This reset behavior is observed experimentally. The experimental data presented by [152, 95], the cells were cultured in media containing EGF. Prior to stimulation the cells were switched to serum and EGF free media for 12 h. Cells treated in this way are able to recover from the EGF contained it their culture media, an, at least after twelve hours return to an EGF responsive state [191].

In the Hornberg model, the dominant mechanism for desensitization and adaptation of the pathway to EGF is endocytosis and degradation of the receptor complex. Opposing this process are constitutive production and degradation reactions for the receptor, which allow the

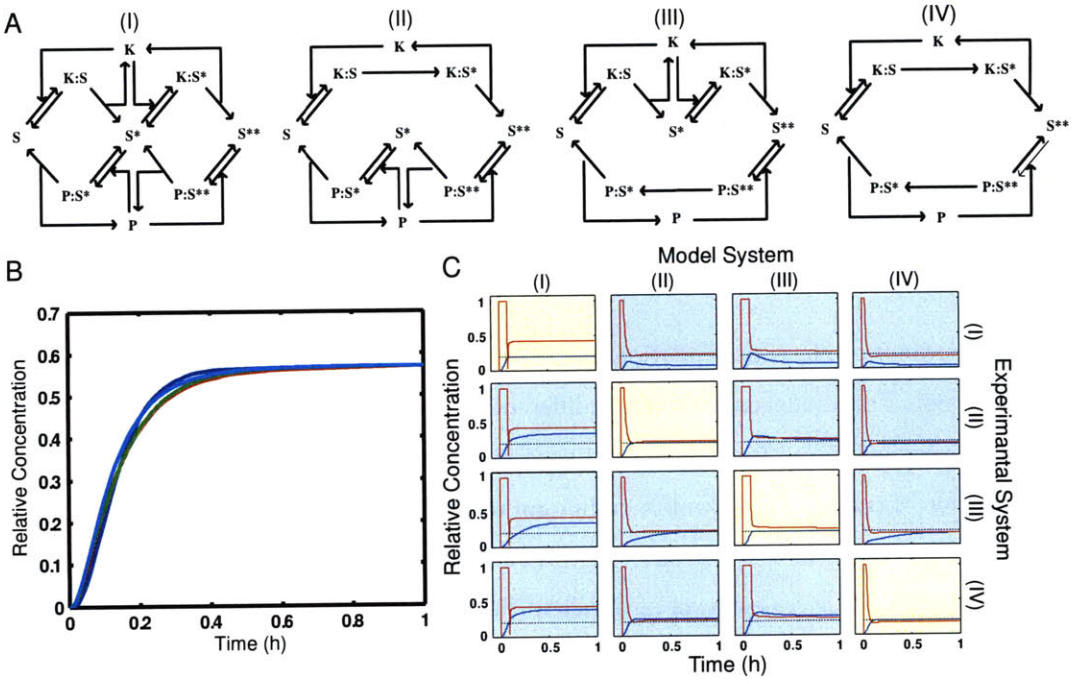


Figure 2-3: Analysis of MAPK Mechanisms. (A) Four alternative MAPK reaction schematics are shown. These correspond to all combinations of processive and distributive kinase and phosphatase mechanisms. Model I is the conical all-distributive mechanism. For each model the input is the concentration of activated kinase (K) and the output is the doubly phosphorylated substrate (S^{**}) (B) All four MAPK models respond in very similar fashion to a step increase in kinase input ($u = 1$). (C) A set of 16 model-selection experiments. Each row is a different experimental system and each column is a different candidate model. Red lines are inputs (activated K), blue lines are the output (S^{**}), and the black dashed line is the design output trajectory. The experiments on the diagonal show that the correct model can control the system. The off-diagonal experiments show that the wrong model does a worse job. This difference can be used to select the correct model.

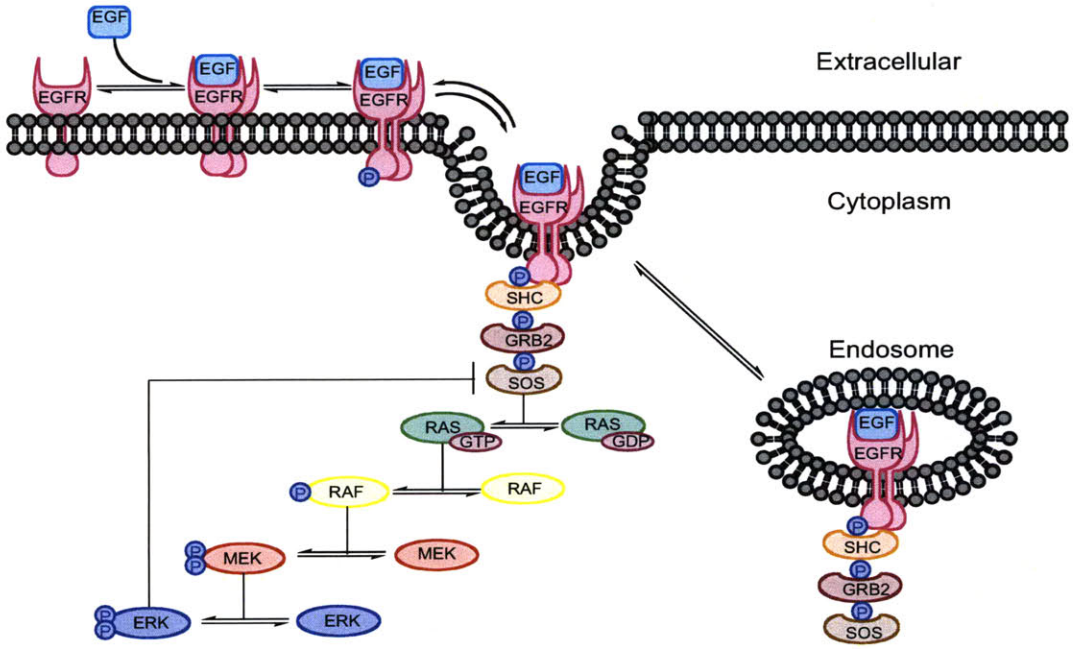


Figure 2-4: Schematic of EGF Induced Signaling. This schematic of EGF induced signaling shows the major steps in EGFR signaling. At the top of the pathway ligand binds to the receptor and induces receptor dimerization and activation. The signal is then transduced through a series of adaptor proteins SHC, GRB2, and SOS, which in turn activates the MAPK kinase cascade RAF, MEK and ERK. There are two negative feedback loops: internalization and degradation of the receptor complex, and ERK deactivating SOS.

receptor level to return back to steady state after stimulation. This same process degrades other proteins in the receptor complex (GAP, GRB2, SOS, and RAS), but the current model does not contain synthesis terms for these proteins. As a result, prolonged stimulation depletes these proteins and prevents the activation of RAF, MEK, and ERK. We added production and degradation reactions analogous to the reactions for the receptor for all of the proteins in the receptor complex. Rate constants were chosen such that the steady-state levels in the absence of stimulation were the same as the initial conditions for the model and the exponential time constant for the approach to steady state was the same as for EGFR.

The second modification to the model was in the RAS-GDP/RAS-GTP cycle. In the Hornberg model, activated receptor is needed to catalyze the recycling of RAS-GTP* (a molecule of RAS-GTP that has already activated a molecule of RAF) that is waiting to be recycled to RAS-GDP. If EGF is removed, RAS can be trapped in the RAS-GTP* form, preventing the system from returning to steady state. We addressed this by adding an additional enzymatic step to recycle RAS-GTP* back to RAS-GDP catalyzed by GAP and parameterized using literature rate constants [113].

With the addition of these new reactions, the modified model returns to its initial conditions after stimulation. For the remaining model parameters (the parameters shared with the original model) we fit the modified model to the original using data from a simulated step-response experiment (Figure 2-5A) constraining them to be within 10% of their original value. Despite the introduction of these new mechanisms and the tight constraints on the parameters, the step responses of the six molecular species modeling those presented in the original paper [152] (Figure 2-5B) are very similar in the original model (blue curves) and the modified model (red curves). The largest difference is in the SHC* time course, which has a very similar shape and varies by at most 11%. While significant, this difference would be very difficult to detect in a standard biological experiment. As such, the modified model is a reasonable alternative to the original model, and it would be hard to reject either mechanism using the step-response data alone.

From this starting point we used our methodology to design an experiment that could distinguish between the current model and the modified model of the EGFR pathway. For each

model we tasked the dynamic optimization controller with driving the concentration of doubly phosphorylated ERK to a constant level of 104 molecules per cell. The input basis set was 25 points linearly spaced over the interval. To model the experimental condition where it is easy to add EGF to the dish of cells but difficult to remove, we implemented a monotonicity constraint. Figure 2-5B shows the inputs designed for each of the two models applied to each system with the resulting ERKpp time courses. Due to the negative feedback loops, both models required a steadily increasing concentration of EGF to maintain a constant level of ERKpp. However, the original model was much more difficult to control; as time progressed increasingly high doses of EGFR were required to maintain a constant output. The modified model required a much gentler increase in EGF concentration to maintain its level and was able to keep the concentration of ERKpp high to the end of the time period. Trial calculations showed that this result was robust to order of magnitude changes in the new rate parameters introduced in the modified model. Applying these two signals in an experiment could be used to distinguish between these two models, as demonstrated by the simulations.

2.4 Discussion

The most common stimulus-response protocol involves applying a step change in one or more input concentrations and following the evolution of one or more downstream molecules. For a linear system, this type of experiment can provide enough information to fully identify the system [135]. However, even simple biochemical systems are nonlinear, and as such there is no *a priori* reason to believe that a step-response experiment will be sufficient to uncover the relevant dynamics of the system and allow for the selection of a unique model. As a result, it is often possible, if not probable, that multiple mechanisms fit the same set of step-response data. We have shown here that using dynamic stimulation can improve stimulus-response experiments. Even in the context of complex pathways with limited numbers of inputs and outputs, experiments can be designed that are capable of distinguishing amongst alternate mechanisms. Moreover, for the EGFR pathway studied, the differences detected were in the middle of the pathway, far from the location of the stimulus or the readouts.

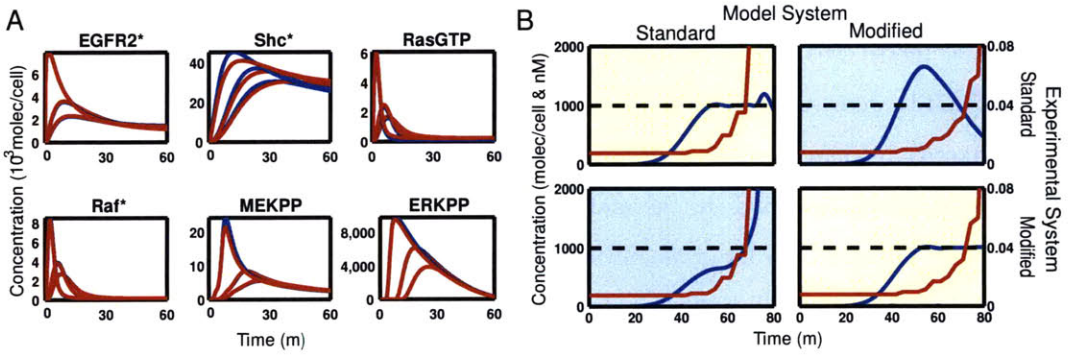


Figure 2-5: Comparison of Step Experiment to Designed Experiment for EGFR Pathway with Original and Modified Models. (A) Both models respond similarly but not identically to a step input in EGF over a range of concentrations. The red lines are the outputs of the modified model and the blue lines are the outputs of the standard model. Often the blue lines are not visible because they are under the red lines. (B) Two designed dynamic stimuli were applied to two models of the EGFR pathway. The red lines show the input (EGF) concentration as a function of time. The blue lines show the output concentrations (ERKpp). The dashed black line shows the target ERKpp concentration. The controller for the standard model is unable to keep the output level high and saturates. In contrast, the modified model requires a more gradual increase in the input and can control the experiment over the course of the entire time course. In both cases the controller based on the wrong model performs worse that the controller based on the correct model.

One possible explanation for the results presented here is that we have stimulated the systems with high-frequency signals, and it is this fact that allows for model discrimination. While the high-frequency content almost certainly plays a part, the fact that the differences between models are observed at low frequency distinguishes our results from other standard test signals. For example, in linear systems it is common to use random or pseudorandom signals to discriminate among models. Figure 2-6 shows such an experiment. While the signal is discriminating, the observed differences are high frequency and would be difficult to distinguish in a standard biological assay, which are usually sampled sparsely in time.

Formulating experiment design as a control problem yielded a relatively straightforward numerical solution, which allowed us to apply our method to large pathway models. While the method does not yield optimal experiments, in the sense of maximizing the least squares error between model ouputs, the results are still of practical benefit and appear sufficient to distinguish amongst model candidates. In the systems studied here, the designed inputs were able to substantially increase the differences observed between competing models when compared to the corresponding step-response experiment. By prescribing the target output trajectory, it should be possible to tailor the experiments to the available measurement methods, thereby achieving the most benefit from existing assays. It is worth noting that in all of the examples presented here, the target function was a constant output concentration. This was chosen for simplicity rather than for any special property of these targets. The problem of the best target function is an interesting one but is beyond the scope of this work. However, in Figure 2-7 we show calculations for the antibody-ligand system using other simple target functions, lines of constant slope, and find that designed inputs based on these signals have similar discriminating power.

In each of the cases presented here, the dynamic stimuli allowed us to select the correct mechanism from a set of plausible candidates. However, it is possible that for a particular system and set of constraints, the algorithms presented here may fail to find a signal that is sufficiently discriminating. In this case a different choice of target function or a more sophisticated optimization approach may yield better results. However, it is worth noting that in the systems studied here both methods were able to find very good solutions in all cases. In

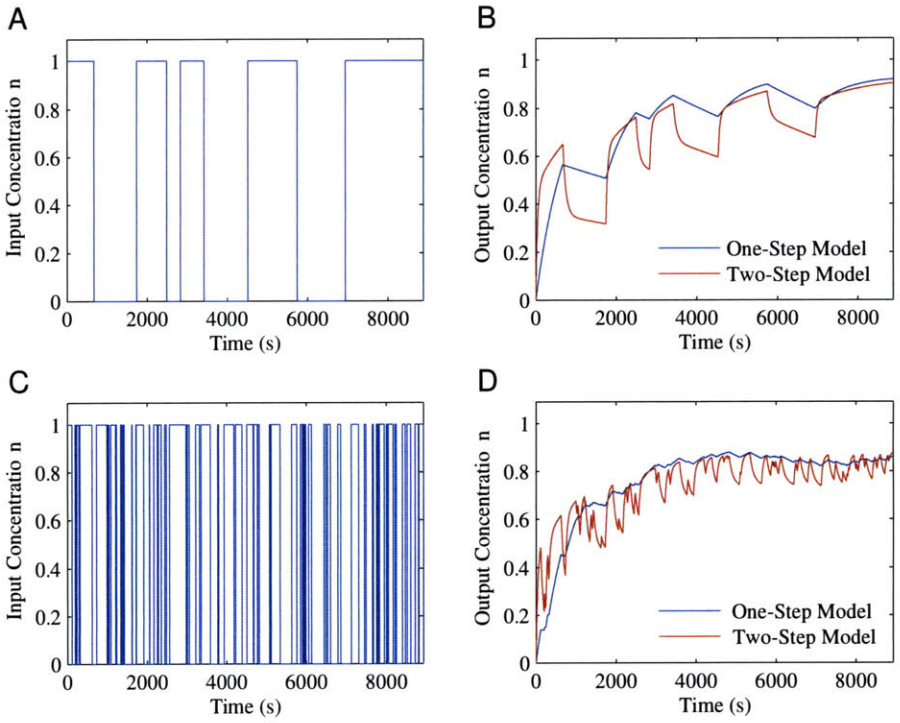


Figure 2-6: Random Pulse Experiment. This figure shows a series of pulses as input to the one-step and two-step antibody ligand models for two different distributions of switching time. In the slower switching time (A) the input signal changes at random with a mean of 900 s, and in the faster switching time (C) the input switches at random with a mean of 90 s. (B) The response of the two models to the slower input shows that the one-step model looks like a smoothed out version of the two-step model. (D) The trend is similar with the faster varying signal.

general, the tangent linear controller was more computationally efficient and yielded smoother signals, whereas the dynamic optimization controller was slower but did not require tuning of parameters such as τ.

One potential limitation of our method comes from our reliance on parameterized models. The accuracy of the parameterizations will affect the quality of the predictions made by the controllers and thus the ability to distinguish between models. To demonstrate this, we generated 100 different parameterizations of the one-step and two-step antibody models and then applied the control signals designed using the nominal parameter set (Figure 2-8). The parameter variation resulted in output trajectories that were quantitatively different from the predicted output trajectories. However, the overall shape of the output trajectories was preserved.

All of the results presented here were in simulation. In practice, experimental error and measurement noise will make it more difficult to distinguish between models. As a result, one may only be able to effectively discard some candidate models, and reduce the pool of hypotheses. However, these experimental challenges also motivate our method, as it has the potential to increase the experimental observability of model differences when compared to a more traditional experiment, such as a step response. Moreover, the fact that potential mechanisms can be evaluated without having to resort to additional inputs or outputs is especially valuable in laboratory experiments, where adding additional inputs or outputs may require significant effort, such as developing new experimental reagents.

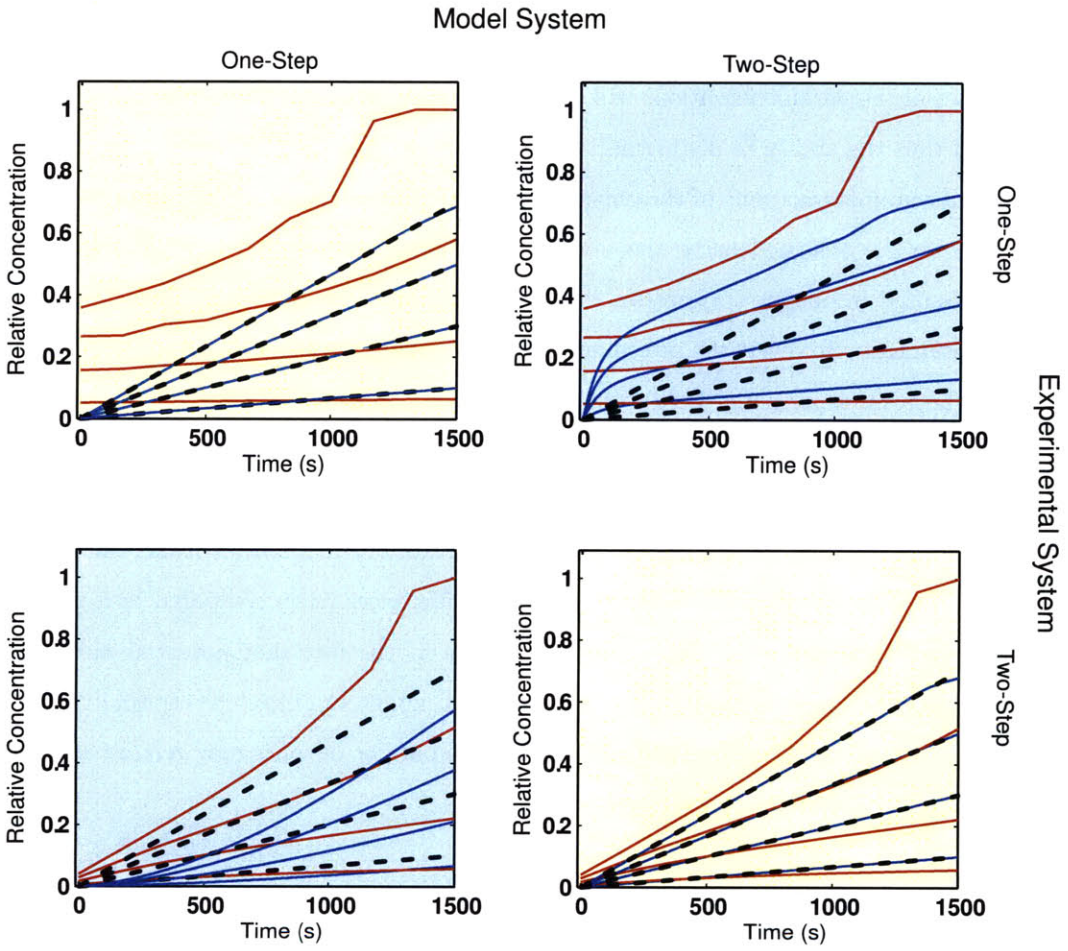


Figure 2-7: Matching Different Target Functions. The dynamic optimization controller was used to design input signals to drive the one-step and two-step antibody ligand models. The target functions (black dashes) are a set of lines of increasing slope. The input is the concentration of free ligand (red) and the output is the amount of complex formed (blue). The figures on the diagonal show that the controller based on the correct model is able to accurately drive the systems to follow the target functions very closely. Whereas, the figures on the off diagonal show that the inputs based on the wrong models cause either large over- or under-shoots of the target function.

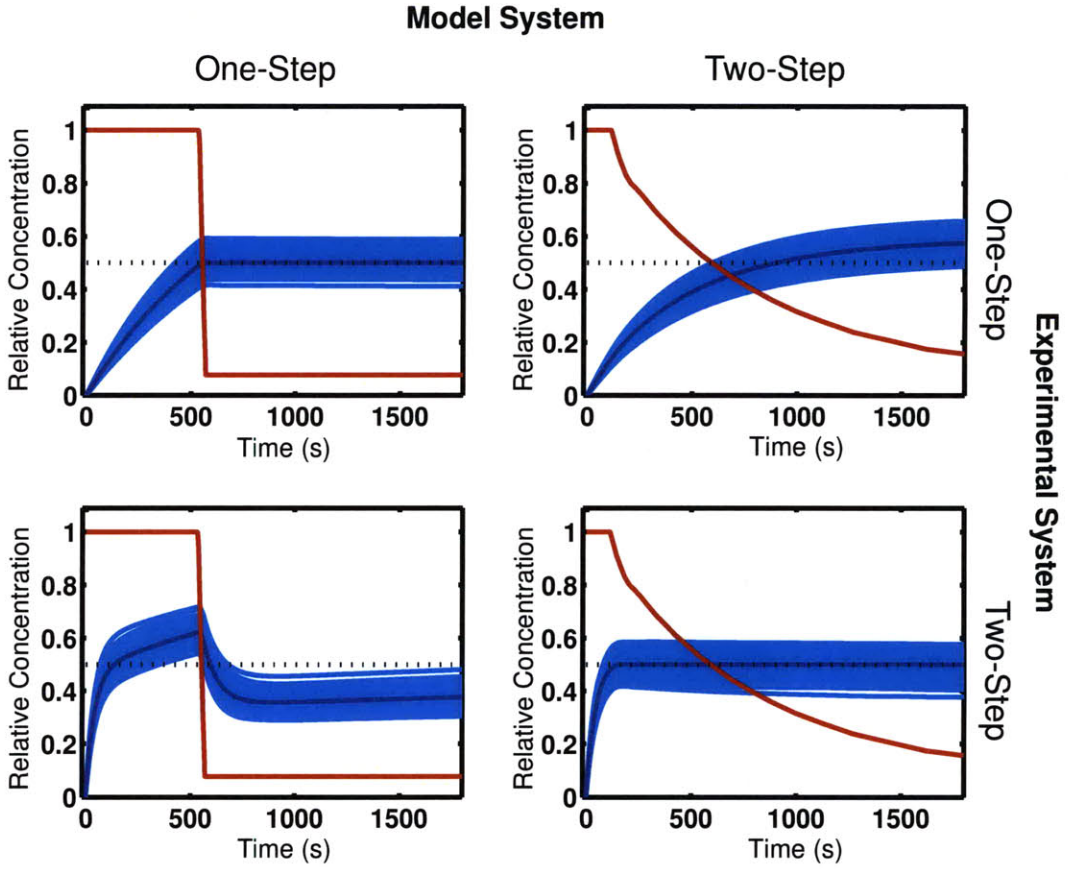


Figure 2-8: The Effect of Parameterization Errors. A dynamic optimization controller was constructed based on a nominal parameterization of the one-step and two-step antibody models. The resulting input signal is shown in red and the output is shown in dark blue. These input stimuli were then applied to a set of 100 different parameterizations of these two models (cyan lines). The parameters were chosen from a log normal distribution with a mean centered on the nominal value and a variance of 10%.

Chapter 3

Using Time Varying Stimulation to Understand the Multiple Feedback Loops in Epidermal Growth Factor Receptor Signaling

3.1 Introduction

Signaling downstream of epidermal growth factor receptor (EGFR) is regulated by a complex network of multiple interconnected negative and positive feedback loops [185, 45, 3, 146, 116, 124]. The full biological implication of these loops is not well understood, but they have been implicated in pattern formation during development [140, 143], carcinogenesis [82], and cell motility [117, 91]. Because of their biological importance, EGFR, as well as many of the downstream kinases, are drug targets [55]. However, the efficacy of these drugs is highly context dependent, and most exhibit significant patient–to–patient variability [138]. The biochemical mechanism of this variability is not well understood [134, 37, 126]. One possible explanation is that, depending on the state of the networks, these negative and positive feedback loops either undermine or enhance the efficacy of a drug [98, 63]. This suggests that understanding the

structures and the regulation of these feedback processes may illuminate the differential pharmacological responses to these drugs, suggest diagnostic biomarkers, and motivate improved clinical regimens including cotherapy strategies.

The EGFR signaling cascade begins with the formation of an active signaling complex. Epidermal growth factor (EGF) binds to the extracellular domain of EGFR, which stabilizes receptor dimers and leads to auto-phosphorylation of several tyrosine residues on the cytoplasmic tail of the receptor. These phosphotyrosine residues serve as docking sites for a complex network of adaptor proteins [90]. The adaptor proteins GRB2 and SOS binding to the receptor, either directly or through Shc, completes the active signaling complex. Once completed, the signaling complex can activate the small GTPase Ras which in turn activates the Raf/MEK/ERK mitogen-activated protein kinase (MAPK) cascade.

Opposing this activating signaling cascade are several strong negative feedback loops. The best characterized mechanism of signal attenuation involves the endocytosis, ubiquitination, and degradation of the receptor [180, 32, 45, 184, 185]. Activated receptors cluster over clathrin-coated pits driving vesicle formation. These vesicles are then transported to the sorting endosome. Some fraction of the receptor and the ligand are recycled to the surface, while a majority of the activated receptors are trafficked to the late endosome and degraded. The fate of the adaptor proteins that are associated with the receptor complexes during the degradation process, however, is less clear. The decision to recycle or degrade the receptor is at least partially regulated by the ubiquitin ligase Cbl [45].

In addition to the endocytic regulation, there are a number of post translational feedback loops involved in silencing the EGFR signal. In one loop, MEK phosphorylates Raf on its 14-3-3σ binding site, which deactivates Raf. In another loop, ERK, a promiscuous kinase which can phosphorylate more than 160 substrates in the cytoplasm and the nucleus [157, 186], phosphorylates SOS, promoting the dissociation of SOS from GRB2 [27]. There is also evidence that ERK can phosphorylate EGFR on threonine–669 which has been shown to lead to decreased receptor internalization [83] and inhibited kinase activity [167]. ERK also activates a number of transcription factors. Recent work by Amit *et al.* [3] has shown that stimulation with a constant saturating dose of EGF leads to the rapid activation of a number of "delayed early

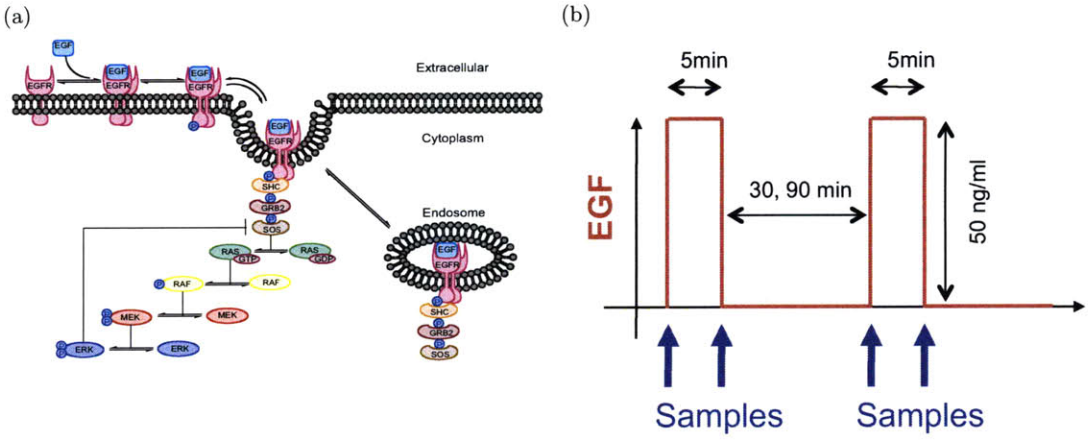


Figure 3-1: Time Varying Stimulation of EGFR Network. (a) shows schematically the EGFR signaling network. (b) shows the EGF time course that was used to stimulate the network.

genes". This set of genes is enriched for negative regulators of the EGFR pathway, including the dual specify protein phosphatases (DUSPs), one of which, DUSP6, is known to be able to dephosphorylate ERK *in vivo* [20]. Amit *et al.* hypothesized that upregulation of the DUSPs formed a negative feedback loop that shut off ERK. To test this hypothesis, cells were pretreated with cycloheximide (a protein synthesis inhibitor) 10 minutes prior to EGF stimulation, with the intent of blocking transcriptional feedback from producing more phosphatase. The result of the cycloheximide treatment was a prolonged phospho-ERK signal [3], which is consistent with their hypothesis. None of the current computational models of the full pathway include this negative feedback loop. However, this mechanism has been observed and modeled by Bluthgen *et al.* in a synthetic system where ERK is activated by the inducible expression of a constitutively active RAS.

In addition to the negative feedback loops, there is a negative feedforward loop controlled by Ras GTPase activating protein (Ras–GAP). When Ras–GAP is bound to Ras, it greatly increases the rate of GTP hydrolysis, which converts Ras from the active form (Ras–GTP) to the inactive form (Ras–GDP). It should be noted that Ras–GAP is a major point of crosstalk between signaling networks and can be activated by many receptors including other Erb Family members [43, 18], TrkA [149] and insulin receptor [21].

To help decipher this complex network of feedback loops, a number of computational models of this pathway have been produced. While these models match the observed data well, they contain a diverse set of mechanisms and are controlled by different principles. One popular model of EGF signaling from the literature was developed by Hornberg and coworkers [86] as a refinement of earlier models [95, 152, 59]. This model includes the formation of an active receptor signaling complex leading to the activation of the MAPK cascade, and contains two negative feedback loops: the endocytosis and degradation of the receptor complex and inhibition of SOS by direct ERK phosphorylation. A very similar model was created by Brightman and Fell [27]. It contains all of the same biophysical processes as the Hornberg model. Despite implementing similar mechanisms, the models operate by very different principles. The Hornberg model is dominated by the endocytosis loop, and removal of the ERK negative feedback loop does not significantly affect the model step response [152]. By comparison, the Brightman and Fell [27] model, is dominated by the ERK to SOS negative feedback loop. Other models developed by Sasagawa *et al.* [149] and Brown *et al.* [31] include an additional negative feedforward loop from EGFR to Ras–GAP. However, the Sasagawa model does not include receptor recycling and the Brown model does not include any receptor trafficking. The Hornberg and Brightman models do not model Ras–GAP explicitly. Instead both models abstract this process and allow the activated receptor complex to catalyze both the activation (Ras–GDP to Ras–GTP) and the deactivation (Ras–GTP to Ras–GDP) reactions.

One possible reason for the lack of consistency between these models is that the experiments used to validate these models are largely insensitive to the weights given to the different mechanisms. The typical experiment for model validation involves stimulating cells in culture with a saturating dose of EGF and then measuring the concentrations of various downstream species at various time points. We call this type of experiment a step response experiment. We have shown previously that such experiments are not particularly well suited to distinguish between different mechanisms [5]. In particular, they are not strongly affected by the details of the negative feedback loop. We believe that the reliance on step response data has led to the ambiguity that exits in models of EGFR signaling. However, in our work we also showed that a time varying stimulus could yield data that was more informative of the detailed mechanism.

In this study, we tested the idea that a time-varying EGF stimulation could reveal more of the subtle regulation of the EGFR network and help elucidate the relative effects of the different feedback loops. Rather than investigate all of the negative feedback loops, we focused on the detailed mechanism of the endocytotic feedback loop. In particular, we tested the mechanistic assumptions of the Hornberg model that lead to the dominance in that model. The stimulus we chose was pulses of EGF separated by different lengths of time (Figure 3-1(b)). The results show that the Hornberg model does not accurately predict the response of the network to this time varying stimulation. We also show that some simple changes that we previously suggested [5] do lead to improved predictions, but that the mechanism proposed is not correct. Based on our findings we suggest some alternative mechanisms that are consistent with this new data, as well as previous results. Although our focus here was on the endocytosis feedback loop, the data presented here seem to be highly sensitive to the impact of of other feedback process. Taken together, these results suggest that this experimental strategy of pulsing the network with EGF combined with specific inhibitors may be an effective strategy to unravel the details of this complex multi-feedback network.

3.2 Results

3.2.1 Modeling Prediction

In the EGF signaling developed by Hornberg and coworkers [86], the endocytosis and degradation of the receptor complex is the dominant negative feedback mechanism. The model assumes that in the late endosome all of the proteins in the receptor complex are degraded. This includes the receptor and all adaptor proteins or kinases bound to the complex. The affect of this assumption is that stimulus with EGF depletes the cytoplasm of adaptor proteins, effectively severing the receptor from the downstream signaling events. While the model includes synthesis terms that replace the degraded receptor, there are no similar terms for the adaptor proteins. As a consequence, this model makes an interesting prediction, that endocytosis and degradation of the receptor complex should make the network insensitive to EGF after an initial stimulation, and that this insensitivity should be persistent. This mechanism has been propa-

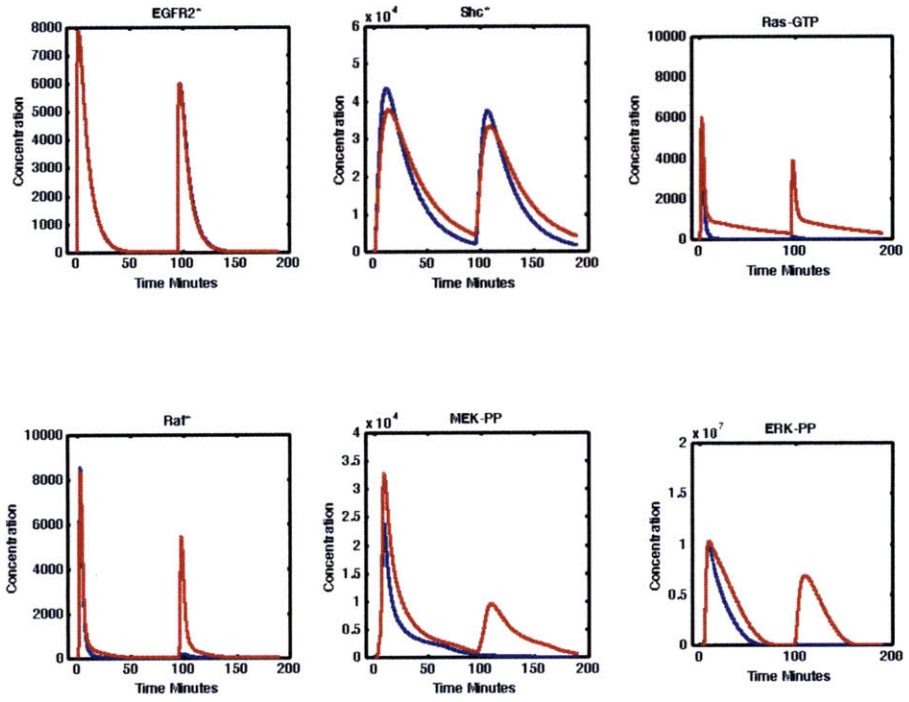


Figure 3-2: Two Pulses of EGF Applied to Models With and Without Adaptor Protein Recycling. Two five–minute pulses of EGF were applied at time zero to the Hornberg model without adaptor protein recycling (blue lines) and to the Hornberg model with adaptor protein recycling added (red lines) The pulses were applied at $t = 0$ min and then again at $t = 95$ min. Both models respond roughly equally to the first and second pulse at the level of EGFR and Shc,. However, the model without adaptor protein does not respond to the second pulse at the level of Ras-GTP or below.

gated to newer versions of this model [43]; however, our intuition about the system is that given sufficient amount of time the network should be able to reset from an transient pulse of EGF. In previous work, we proposed adding adaptor protein recycling via constitutive synthesis and degradation [5]. Like the Hornberg model, the model with recycling is still strongly regulated by endocytosis and degradation. Unlike the Hornberg model, the model with recycling is able to recover to its pre-stimulated state after transient EGF stimulation. Interestingly, the addition of these reactions did not significantly change the step response behavior of the network. However, it was shown that the two models responded very differently to a time varying stimulus that consisted of a ramping dose of EGF. The stimulation described in that work would be difficult to realize experimentally. Instead, a simpler time varying stimulus was used.

In this study we stimulated cells with five–minute pulses of EGF (50 ng/mL) separated by a variable waiting time were applied to cells in culture. The idea behind this stimulus is that the first pulse activates the network, including all of the feedback loops. The second pulse then, probes the state of the network at a later time. However, we expect that as the response to the first pulse evolves, different feedback mechanisms will increase or decrease in importance. By comparing the response of the network the second pulse delivered at various times we hoped to be able to detect the timescale and relative importance of the various feedback mechanisms. In particular, we were interested to see if we waited long enough after the first pulse, would the network recover and become sensitive to EGF again. To accesses the potential for this type of experiment we simulated applying two pulses separated by 90 minute pulse waiting time (Figure 3-2). At the head of the pathway, the models with recycling (red lines) and without recycling (blue lines) respond similarly to the first and the second pulse. However, the model without recycling predicts that the signals downstream of the adaptor proteins will not be strongly affected by the second EGF pulse. In contrast, the model with recycling predicts that the response to the second pulse will be similar to the response to the first pulse. This difference between these two models is the greatest in the ppERK response.

To test this hypothesis, we conducted series of experiments in which cells were stimulated with two 5–minute pulses of EGF. The experimental system we choose was 184A1 human mammary epithelial cells (HMEC) [163]. This cell line has been well studied for its response

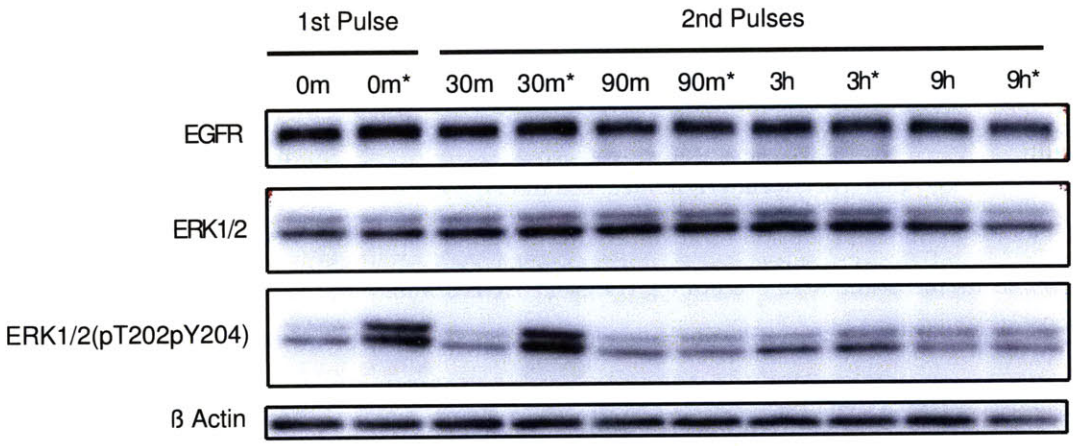


Figure 3-3: Varying the Time Between EGF Pulses. Cells were stimulated with two 5–minute pulses of EGF (50 ng/mL). The time between pulses was varied (30 min, 90 min, 3 hr, and 9 hr). The lanes are labeled with the time between the two pulses, and 0 indicates the first pulse. As the protocol for the first pulse is the same, it was not repeated for each pulse waiting time. Cell lysates were collected just before and just after each pulse. The lysates that were taken after the pulse are indicated with a star.

to EGF [191]. The stimulation protocol is described in detail in the methods section, but, in brief, at time 0 EGF was added to plates of HMEC cells. After 5 minutes, the EGF containing media was removed, the plates were washed quickly with warm PBS, and then fresh serum free media was added. After the pulse waiting period (variable), EGF was again added to the plates and the cells were allowed to incubate for an additional 5 min. Cell lysates were collected just before and just after the first pulse and just before and just after the second pulse.

3.2.2 The EGFR Network Can Reset After a Pulse of EGF

To establish how long the network would take to reset (if at all), we performed an initial survey experiment with four different pulse waiting times: 30 min, 90 min, 3 hr, 9 hr. The output of the network, doubly phosphorylated ERK (ppErk), was measured just before and just after each pulse by western blot. As the first pulse is identical for the four pulse waiting times, these measurements were not repeated. The results of this experiment are shown in Figure 3-3.

3.2.3 The EGFR Network Desensitizes to EGF Possibly Due to Cross Talk From the Insulin Pathway

Our expectation from the modeling results was that the longer the time between pulses of EGF the stronger the ppERK response would be in response to the second pulse. This is due to the network having more time to replace the degraded proteins. However, what we saw from this study was just the opposite. If the pulse waiting time was 30 min then the responses to the first pulse and the second pulse were about equal. However, if the pulse waiting time was 90 min, 3 hr, or 9 hr then there was almost no response to the second pulse. These results suggest that there is some kind of desensitization that happens after the first pulse but does not become apparent until between 30 and 90 minutes.

One possible explanation for this apparent desensitization is that the initial pulse of EGF activated an endocytic process that over time depleted the surface of EGFR. However, the anti-EGFR western blots indicate that the total amount of EGFR is roughly constant over all of the time courses. This measurement cannot distinguish EGFR on the extra cellular membrane from internalized EGFR. However, the fact that the level of EGFR is roughly constant over 9 hours, and the fact that after the first five minutes there is no EGF stimulus to drive endocytosis, suggests that this was not responsible for the observed desensitization.

Another possible explanation for this result was that 12 hr serum starvation had induced some kind of cell death or senescence. To test this, we repeated the pulse experiments, this time adding back in the growth media components to the pre-treatment media (Figure 3-4). As before, the media was changed 12 hours prior to EGF stimulation. In one experiment, the pre-treatment media was supplemented with 1% fetal bovine serum (FBS). In another experiment the pre-treatment media was supplemented or 1% FBS and 25 ng/mL EGF. The serum free media results from before are repeated here for comparison. The most striking difference between these two sets is the high background phosphorylation of ERK, presumably due to the FBS and EGF in the media. However, the cells pretreated with media containing 1% FBS behave in a similar fashion to the serum starved cells. The background in the FBS and EGF pretreated cells is too high to draw a firm conclusion. However, it is worth noting

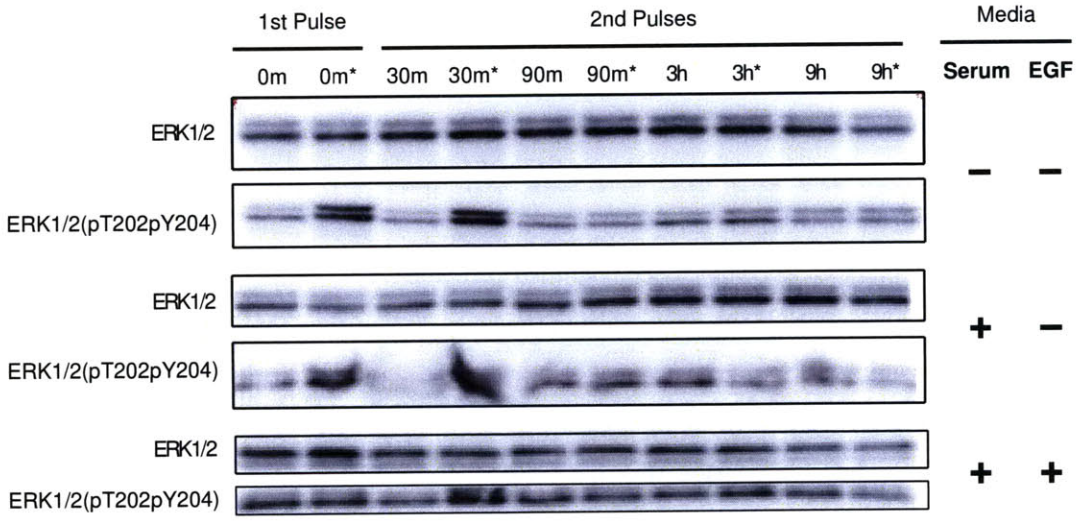


Figure 3-4: Double Pulse Experiments Without Serum Starving. The double pulse experiments were repeated under different pre-stimulus media conditions. Cells were stimulated with two 5–minute pulses of EGF (50 ng/mL). The time between pulses was varied (30 min, 90 min, 3 hr, and 9 hr). The lanes are labeled with the time between the two pulses, and 0 indicates the first pulse. As the protocol for the first pulse is the same, it was not repeated for each pulse waiting time. Cell lysates were collected just before and just after each pulse. The lysates that were taken after the pulse are indicated with a star. The media conditions are indicated to the right of the corresponding blots. In all three cases, the media was changed 12 hours prior to stimulation. -/- indicates that the media was serum free media. + serum indicates 1% fetal bovine serum was added and + EGF indicates that 25 ng/mL of EGF was added to the media.

that the darkest band is observed when using a 30 min pulse waiting time, just as in the serum free experiments. These results suggest that the desensitization to EGF is not the result of the 12 hr serum starving.

Another possible explanation for this result is that washing the cells with PBS and then changing the media produced mechanical stresses on the cells that lead to the desensitization of EGF. One possible mechanism for this would be signaling through the integrins. We repeated the first set of experiments but either skipped PBS wash and very gently changed the media, or washed twice with PBS swirling the plates during each wash. Both treatments produced similar results to the first data set (data not shown).

Another possible explanation for this result is that some component of the serum free media added after the first pulse was inducing the desensitization. In particular, the media contained

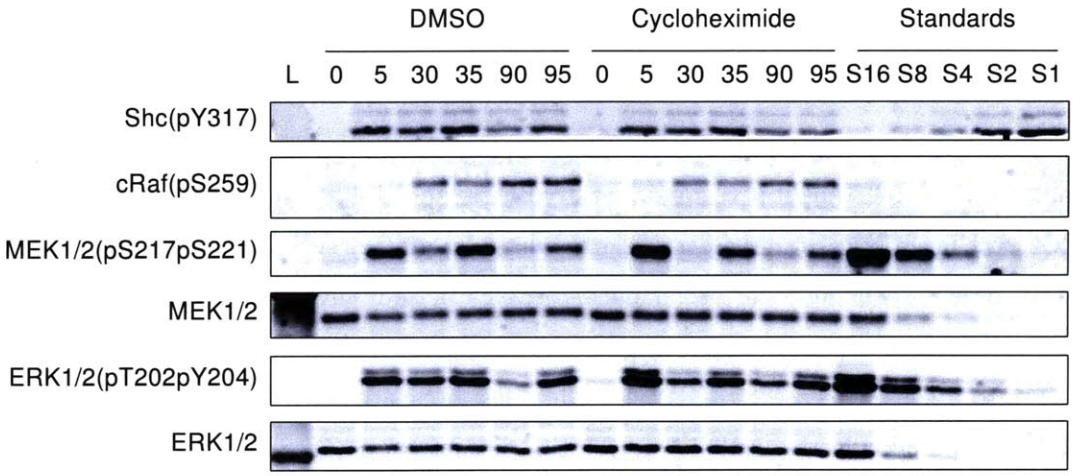


Figure 3-5: Stimulating HMEC cells with pulses of EGF. HMEC cells were stimulated with two 5–minute pulses of EGF (50 ng/mL). Lysates were taken just before and just after each pulse. In one experiment, the pulses were applied at 0 minutes and 35 minutes (lanes 0,5,30,35). In a second experiment, pulses were applied at 0 minutes and 95 minutes (lanes 0,5, 90, 95). As the protocol for the first pulse is the same in the two experiments, the lysates were not duplicated. The full experiment was repeated twice: without drug treatment (DMSO), and with a protein synthesis inhibitor (cycloheximide).

20 μg/mL insulin. We repeated the double pulse experiment this time with just two pulse waiting times (30 min, and 90 min). We used the same serum free media as in the first experiments but with the insulin removed. The results are shown in Figure 3-5. Unlike all of the previous experiments, the ppERK response after a pulse at either 30 or 90 minutes is roughly equal to the response to the first pulse. In addition, the level of ppERK 30 minutes after the first pulse (before a second pulse) is much higher than before. This suggests that insulin was playing a major role in turning off the ppERK signal even at 30 minutes. One possible mechanism for this is insulin induced activation of Ras-GAP. However, the fact that even without insulin ppERK is down significantly 90 minutes after the first pulse indicates that there are other mechanisms working as well.

3.2.4 Resetting the Network Does Not Require New Protein Translation

The time course of the results without insulin agree well with the computational model with recycling. In the computational model, the network is able to reset after a transient EGF stimulus due to the constitutive (and fast) synthesis and degradation of adaptor proteins. If this mechanism is correct then transiently inhibiting protein synthesis should prevent the network from responding to a second pulse of EGF. In addition, other studies have shown that cycloheximide treatment results in an stronger and more prolonged ppERK signal [3] in response to a constant dose of EGF. This effect has been attributed to a rapid turnover of phosphatases especially the dual specify protein phosphatase 6 (DUSP6) [20]. Therefore, we expected cycloheximide treated cells to exhibit a prolonged response to the first EGF pulse, and a diminished response to the second pulse, at least at the level of ppERK.

However, the results in Figure 3-5 confounded this expectation. In the cycloheximide treatments the ppERK response to the first pulse is slightly elevated at 5 minutes but significantly diminished 30 minutes after the pulse. Moreover, the cycloheximide treated cells responded just as strongly to second pulses of EGF after 30 minutes, and an only slightly weaker response to the 90 minute pulse when compared to the non-drug treated cells. Interestingly, the ppERK levels are elevated 90 minutes after the first pulse when compared to the non-drug treated cells, consistent with the idea of a DUSP6 negative feedback loop.

These results show that the network is, in fact, able to reset after an initial pulse of EGF and the mechanism does not require the new synthesis of proteins. In addition, the negative transcriptional loops that have been described do not seem to be necessary to shut off the EGF signaling after a five-minute pulse.

3.2.5 Phosphorylation of C-Raf on Serine-259

The MEK response is somewhat less after the 90 minute pulse. This may be due to deactivation of C-Raf by phosphorylation at S259 which is the 14-3-3σ binding site [162]. Phosphorylation at the site is higher 30 minutes after the first pulse and significantly higher 90 minutes after the first pulse. However, the fact that there is still strong activation of MEK indicates there

is at least some active C-Raf. An interesting followup experiment would be to apply a second pulse later than 90 minutes to see if at later times inhibition of C-Raf builds to a level to be sufficient to block the activation of MEK.

Another unexpected result is that phosphorylation at pS259 is diminished in the cycloheximide treatment. It is unclear what mechanism would lead to this effect, especially since MEK is strongly activated under these conditions.

3.3 Discussion and Future Directions

Like many cell signaling models, current computational models of EGFR signaling were developed to describe steady state and step response data [43, 95, 152, 59, 21, 86], and less care is taken model how the systems recover after the initial stimulus. We have shown here that the EGFR signaling network can indeed reset after a pulse of EGF. The combined experimental and computational results presented here indicate that existing models of the EGFR signaling are missing a mechanism for the network to reset and to become sensitive to EGF again. However, we have also shown that the mechanism we previously proposed for the network to reset, the synthesis of new receptor and adaptor proteins, does not seem to be the primary mechanism. One possible alternative mechanism that is consistent with this data is that the receptor complex dissociates in the endosome and releases the adaptor proteins. This would replenish the cytoplasmic pools of adaptors without needing new protein synthesis. However, additional experiments are needed to test this hypothesis. In particular, it is unclear from this data if the network could reset after a longer pulse EGF without new protein synthesis.

In addition, the results presented here suggest that even a small amount of insulin may be sufficient to modulate the ERK response to EGF stimulation. A recent model of EGF and insulin crosstalk describes some of the mechanisms by which this could happen [21]. In the model, insulin receptor activates Ras–GAP which in turn inhibits Ras. However, the model also describes the activation of IRS which binds to GRB2 and SOS leading to the activation of Ras. The study found that cotreament of insulin and EGF lead to increased ERK activation. The results presented here suggest that if insulin had been administered first it would have

suppressed ERK activation. Very similar types of crosstalk have been implemented in other RTK models. In the Sasagawa EGF/NGF model, Ras–GAP can be activated by both EGFR and TrkA [149]. In the Birtwistle model, various Erb dimers activate Ras–GAP [18].

An additional feedback mechanism that could be added to these models is the phosphorylation of C-Raf by MEK on S259. These results show that phosphorylation at this site is slowly building over 90 minutes. An additional followup study to this could vary the pulse duration and waiting times to see if phosphorylation at this site ever built up to a level where C-Raf could no longer effectively activate MEK. In addition, it would be interesting to see if the rate of S259 phosphorylation depended on the amplitude or duration of the EGF pulse. All of the models presented in this work are intended to capture acute signaling on the timescale of 0–120 min. Adding in effects such as this S259 phosphorylation may be necessary for the models to capture the longer time dynamics of the system.

In one sense, these results are frustrating to a modeler. They seem to confound the existing models and mechanism. However, in another sense these results are encouraging to future modeling efforts in so far as they suggest that it is possible to construct experiments that are sensitive to the details of the various feedback loops. In particular, the strategy put forth in this work of combining pulsatile simulation with inhibitors that can cut particular feedback processes may be be a very powerful technique for understanding these complex multi-feedback regulated networks.

3.4 Materials and Methods

3.4.1 Cell Culture

184A1 parental human mammary epithelial cells [163] were a kind gift from Martha Stampfer (Lawrence Berkeley Laboratory, Berkeley, CA) and were maintained in MEBM (Lonza Catalog No. CC-3150) supplemented with 35 μg/mL bovine pituitary extract (BPE), 12.5 ng/mL epidermal growth factor (EGF), 1 μg/mL insulin, 28 μM hydrocortisone, 5 μg/mL transferin, 1 ng/mL cholera toxin, 100 μg/mol penicillin and streptomycin, 1% (v/v) fetal bovine serum (FBS) (all from Sigma). Cells were plated in 10-cm plates ($\approx 2 \times 10^6$) and allowed to grow until

80% confluence was reached. The cells were then washed with PBS and serum starved for 12 hours in serum-free media (MEBM without EGF, BPE, or FBS). In some of the experiments, the "serum-free" media was made with FBS, or FBS and EGF. In other experiments, the serum-free media was made without EGF, BPE, FBS, or insulin.

3.4.2 Stimulation

Cells were pre-treated for ten minutes with either cycloheximide (Sigma C4859) or DMSO (Sigma D2438) diluted in PBS. Cells were then stimulated by the addition of epidermal growth factor (50 ng/mL). After 5 minutes the media was removed and replaced with fresh serum-free media containing either cycloheximide or DMSO. After either 30 or 90 minutes, the cells were stimulated a second time with epidermal growth factor (50 ng/mL) for 5 minutes. Cells were lysed at 0 and 5 minutes post stimulation and then again at either 35 and 40 or 95 and 100 minutes, depending on the pulse times. Cells were lysed in 250 μL RIPA lysis and extraction buffer (Thermo 89900) with Halt protease and phosphatase inhibitor (Thermo 78440) and 0.5 mM EDTA.

3.4.3 Western Blot Analysis

30 g of protein from each sample was mixed with 4$\times$ sample buffer (250 mM Tris-HCl pH 6.8, 8% SDS, 40% glycerol, 0.04% bromophenol blue, and 400 mM dithiothreitol) and boiled for 5 min. Proteins were separated by SDS-PAGE and transferred onto nitrocellulose membranes. After blocking for 1 hour with Odyssey Blocking Buffer (Licor) at room temperature, membranes were incubated overnight at 4 C in primary antibody, washed three times for 5 min in TBS-Tween 20 (20 mM Tris-HCl, pH 7.5, 137 mM NaCl, and 0.3% Tween 20), incubated for 1 hour at room temperature in secondary antibody (dilution 1:10000 IRDye 800CW and 700CW conjugated goat anti-rabbit) (Licor), and finally washed three times for 5 min with TBS-Tween 20. Blots were imaged with an Odyssey infrared imaging system (Licor). Primary antibodies used were from Cell Signaling Technologies: MEK1/2 (4694), EGF Receptor (2239), Phospho-c-Raf(Ser289/296/301) (9431), Phospho-Shc (Tyr239/240) (2434), Phospho-Erk1/2

(Thr202/Tyr204) (4377), Erk1/2(9107).

Chapter 4

Sloppy Models, Parameter Uncertainty, and the Role of Experimental Design

4.1 Introduction

One of the goals of systems biology is the construction of computational models that can accurately predict the response of a biological system to novel stimuli [97, 101, 181]. Such models serve to encapsulate our current understanding of biological systems, can indicate gaps in that understanding, and have the potential to provide a basis for the rational design of experiments [42, 5, 52], clinical interventions [102, 130, 63], and synthetic biological systems [54]. There are many varieties of computational models ranging from abstracted data-driven models to highly detailed molecular-mechanics ones. In this report we focus on the popular class of ordinary differential equation (ODE) models [50, 57, 64, 80, 87, 92, 139, 152] typically used to describe systems at the biochemical and pharmacokinetic level, but they are also appropriate at more abstract levels. Constructing an ODE model is comprised of writing kinetic rate equations that describe the time rate of change of the various chemical species (representing the model topology), and determining the unknown parameters in those equations (typically rate constants

and initial concentrations). Unknown parameters are estimated from a variety of data that often includes time-course measurements of concentration or activity. In this study, we have focused on the estimation of parameters, which is often referred to as model calibration. Using computational modeling and experimental design methodology, we have found that selection of a set of experiments whose members provide complementary information can lead to efficient model calibration.

In this work we formulate the model calibration problem as a nonlinear least squares optimization problem, where the goal is to find the set of parameters that minimizes the fit metric [78],

$$\chi^2(\mathbf{p}) = \frac{1}{2n_s n_c} \sum_s \sum_c \frac{1}{T_c} \int_0^{T_c} \frac{[y_{s,c}(\mathbf{p},t) - y_{s,c}(\mathbf{p}^*,t)]^2}{\sigma^2_{s,c}(\mathbf{p}^*,t)} dt \tag{4.1}$$

where n_c is the number of experimental conditions, n_s is the number of species for which measurements are available, the indices c and s run over the conditions and species, respectively, T_c is the length of the time course for condition c , $y_{s,c}(\mathbf{p},t)$ is the model output for species s and condition c at time t with parameter set $\mathbf{p}$, $y_{s,c}(\mathbf{p}^*,t)$ is the corresponding output for the true model parameterization, and $\sigma^2_{s,c}(\mathbf{p}^*,t)$ is a weighting factor that is often taken as proportional to the uncertainty of the experimental measurement.

There is a significant amount of work devoted to how best to solve this optimization problem for biological models [10, 89, 125, 144]. However, in any experimental system, there will always be uncertainty in the data, which means there will be some range of parameter values that, while not optimal, cannot be excluded based on the data. Given a maximum acceptable fitting error ($\chi^2_{\max}$), the calibration problem becomes that of finding all parameters such that the error is less than this threshold. In a neighborhood around the optimum parameterization ($\mathbf{p}^*$) the least squares cost function can be approximated by its Taylor series expansion:

$$[\chi^2_{\max} - \chi^2(\mathbf{p}^*)] \geq \frac{1}{2} [\log(\mathbf{p}) - \log(\mathbf{p}^*)]^T \mathbf{H} [\log(\mathbf{p} - \log \mathbf{p}^*)] . \tag{4.2}$$

Equation 4.2 describes an n_p-dimensional ellipsoid in log-parameter space (n_p being the number of fitted parameters), where all of the parameterizations inside the ellipsoid are feasible. Here

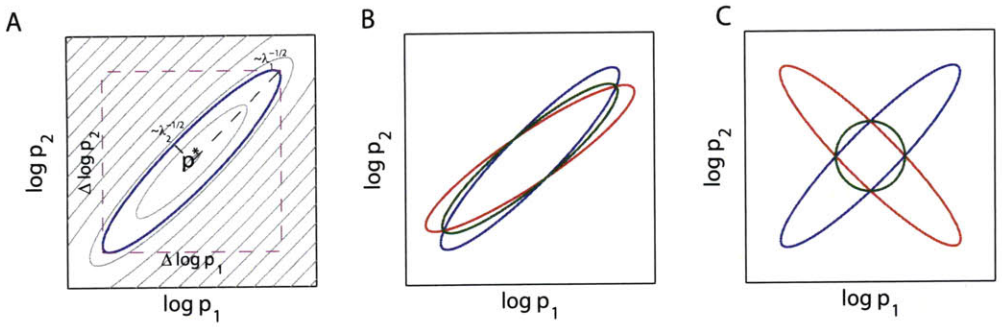


Figure 4-1: Schematic View of Combining Experiments. (A) Uncertainty ellipse for a simple two-parameter system. The parameters inside the ellipse are feasible. The major and minor axis of the ellipse are proportional to $\lambda_1^{-1/2}$ and $\lambda_2^{-1/2}$. The grey contours are lines of equal objective function value. The bounding box (dashed magenta lines) show the single parameter errors. The length and width of the bounding box is the range of values for the individual parameters. (B) Two non-complementary experiments (C) Two complementary experiments. When two experiments (blue and green ellipses) are combined the resulting parameter estimate (red ellipses) can improved as in the case of panel C or not as in the case of panel B depending on the complementarily of the component experiments.

we define $\log \mathbf{p} = [\log(p_1), \log(p_2), \ldots, \log(p_{n_p})]^T$. The size and shape of this ellipsoid describe the multidimensional parameter uncertainty. For example, the longest principal axis of the ellipsoid corresponds to the parameter direction (that is, the linear combination of parameters) with the worst error. Likewise, the axis-parallel bounding box defines the error range for individual parameters. Figure 4-1 shows an example for a two-dimensional system. An important distinction illustrated in Figure 4-1A that is some parameter directions can have very small uncertainty, while the individual parameters can be quite uncertain [78]. For example, forward and reverse kinetic constants for a binding reaction may be poorly constrained, yet the equilibrium constant (given by their ratio), can be well defined.

$\mathbf{H}$, the matrix of second derivatives of the fit metric is known as the Hessian, where $H_{i,j}$ is the derivative of χ^2 with respect to $\log(p_i)$ and $\log(p_j)$.

$$H_{i,j}(\mathbf{p}) = \frac{1}{n_s n_c T_c} \sum_{s,c} \int_0^{T_c} \frac{\frac{dy_{s,c}(\mathbf{p},t)}{d\log(p_i)} \frac{dy_{s,c}(\mathbf{p},t)}{d\log(p_j)} + [y_{s,c}(\mathbf{p},t) - y_{s,c}(\mathbf{p}^*,t)] \frac{d^2 y_{s,c}(\mathbf{p},t)}{d\log(p_i) d\log(p_j)}}{\sigma_{s,c}^2(\mathbf{p}^*,t)} dt \qquad (4.3)$$

The derivative is taken with respect to the natural logarithm of p_i to obtain a relative uncertainty. We can dissect the parameter uncertainty in terms of the eigenvalues λ_i and eigenvectors v_i of $\mathbf{H}$. The eigenvectors form a natural coordinate system for the ellipsoid, pointing along the principal axes. The lengths of the axes are proportional to $\lambda_i^{-1/2}$, meaning that smaller eigenvalues correspond to larger relative parameter error [10].

Recently Gutenkunst et al. [78] examined parameter uncertainty for 17 models in the EMBL BioModels Database [106]. In their study, the authors assumed noise-free measurements of every model species sampled continuously in time. The study found that the eigenvalues of the Hessian spanned a large range ($> 10^6$). From result this they suggested that, while it may be possible to estimate some parameters from system-wide data, in practice it would be difficult or impossible to estimate most of the parameters even from an unrealistically high-quality data set [38, 78, 179]. Moreover, they pointed out that due to the high eccentricity and skewness of the uncertainly ellipses in parameter space, system-wide data can define system behavior better than independent measurements of each parameter and may also produce better predictions in some circumstances.

Here we extend the previous work by more fully considering the effect of experimental perturbations on the parameter estimation problem and use experimental design to probe for particularly effective perturbation experiments. Equation 4.1 shows that the value of the fit metric depends on both the model and the set of experimental conditions; in the current work we use effectively continuous-time data, but most experiments require the selection of discrete time points for measurements to be taken [52]. It is well established in the systems biology literature that optimal experimental design can have an impact on the parameter estimation problem for a single experiment [7, 8, 38, 56, 103]. For example, work by Faller et al. has shown for a small model of a mitogen activated protein kinase (MAPK) cascade that the application of time-varying stimulation significantly improved the parameter estimation problem [56]. Essentially this corresponds to finding the time-varying input signal that gives the best shaped error ellipsoid.

In this work, we apply a related approach and examine the extent to which multiple complementary experiments can be combined to improve the overall parameter estimation problem.

Figure 4-1B,C shows the result of combining data from two separate experiments. The parameter estimates from the individual data sets (blue and green ellipses) tightly constrain one parameter direction and weakly constrain the other. In Figure 4-1B, the weakly constrained parameter directions are very similar, so the parameter estimates from the combined data set are about the same as the estimates from the individual experiments (red ellipse); by contrast, in Figure 4-1C the experiments are complementary and together dramatically constrain the parameter estimates.

Because complementary experiments can constrain parameter estimation space, we have developed an approach to identify sets of complementary experiments to optimally minimize parameter uncertainty in a large pathway model of signaling in response to EGF and NGF [31]. We have selected this model so that our results may be directly compared to the previously published analysis of this model performed by Gutenkunst et al. [78]. For consistency, where possible we have used their methods and formalisms. In selecting sets of complementary experiments, we have explored a pallet of candidate experiments consisting of overexpression or knockdown of single and multiple genes combined with different doses of EGF and NGF, either alone or in combination.

Computational experimental design methods determined all 48 free parameters to within 10% of their value using just five complementary experiments. Selection of *complementary* experiments was essential, as the same level of model calibration could not be achieved with arbitrary experiments or even with a larger number of "highly informative" experiments. Moreover, we argue that predictions that are sensitive to information complementary to that used to parameterize a model could be significantly in error. Experimental design methods can provide sufficient coverage for all parameter directions and thus guide model calibration for a given topology to maximize predictive accuracy. As systems biology models are applied to target identification and clinical trial design, the use of experimental design approaches to improve model prediction quality could be of crucial importance.

4.2 Results

Experimental design and computer simulations were applied in tandem to select a collection of experiments that together could most directly establish each of the rate constant parameters for the EGF/NGF signaling pathway modeled here. To define all 48 rate parameters, experimental design procedures must select a set of experiments that together exercise the model in complementary and sufficiently different ways, rather than simply choosing multiple different experiments that exercise the model in similar ways. For this work we chose a palette of experimental perturbations consisting of stimulation with EGF (10^7, 10^5, 10^3, 10, or 0 molecules/cell–volume) or NGF (4.52×10^7, 4.52×10^5, 4.52×10^3, 45.2, or 0 molecules/cell–volume) individually, or combined treatment with both ligands. We supplemented this choice of ligand stimulation with a panel of experiments in which protein expression levels could be modulated by 100-fold overexpression or knockdown for individual proteins in the network. We then constructed candidate experiments from the combination of ligand choice and protein expression level changes; specifically, each experiment was allowed to comprise one stimulation pattern and up to three simultaneous changes in protein expression level. This experimental set- up resulted in a trial perturbation set of 164,500 individual computational experiments.

4.2.1 All Parameters Can Be Determined to High Accuracy

Each individual experiment in the trial perturbation set was evaluated and the number of rate parameter directions determined to within 10% of their nominal value was recorded. The best individual experiment in this set defined only 29 of the 48 rate parameter directions to this high level of accuracy. In order to improve on this result, each single experiment was re-evaluated to determine how many new rate parameter directions could be defined to within 10% of their nominal value when combined with the best experiment. In this manner, a greedy algorithm was applied to sequentially select sets of experiments based on their ability to generate tighter bounds on rate constant estimation. The results of this greedy algorithm are shown in Figure 4-2A. The parameter uncertainties are expressed as an eigenspectrum for each combination of single experiments, with increasingly larger experimental sets displayed

along the abscissa and eigenvalues displayed along the ordinate. The horizontal dashed line indicates the 10% error level, and the number of eigenvalues above the dashed line represents the number of parameter directions determined to the 10% error level. In Figure 4-2D, the number of parameters estimated to the 10% error level is shown as a function of the number of experiments within the experimental set. It is important to note that, by properly choosing the correct combination of experiments, only five total experiments are sufficient to determine all 48 directions (and, indeed, all 48 actual parameters) to within 10% accuracy. This result indicates that parameter uncertainty, rather than being inherent to biological models, can be progressively reduced by perturbation experiments.

The five experiments determined here to elucidate all 48 parameters are listed in Table 4.1. The selected experiments include a tendency for dual stimulation with both EGF and NGF, combined with multiple protein-expression changes, and a preference for overexpression as opposed to knockdown of given proteins. Interestingly, the experiments do not appear to systematically explore all regions of the perturbation space. For instance, four of the five experiments have low dose of EGF stimulation, and three of the five experiments have high dose of NGF stimulation. Additionally, NGF stimulation occurs in the absence of EGF stimulation in one experiment, but EGF stimulation alone is never utilized. The combination of experiments and the manner in which they explore different aspects of the model in order to adequately define all parameters is not readily apparent from selected experiments. However, as will be discussed below, there are some general trends and insights to be gained from the analysis of the results.

Figure 4-7 shows the location of parameters in the model determined after each new experiment was added in sequence to the set. The thickness of each reaction arrow in the figure indicates whether 0 (thin arrow), 1 (medium arrow), or 2 (thick arrow) of the parameters associated with the arrow are known to within 10% at that point in the experimental sequence. Note that parameters closer to the top of the pathway tend to be determined first, while parameters toward the bottom of the pathway, further from the application of ligand stimulation, tend to be determined only after multiple experiments are combined.

The selection of complementary experiments resulting from our experimental design procedure is non-trivial. For example, if experiments were added sequentially based on their ability

Exp	EGF (mol./cell)	NGF (mol./cell)	Over Expressed	Knocked Down
1	1.00×10^5	4.56×10^7	Sos, Ras, C3G	
2	1.00×10^1	4.56×10^1	Mek, Erk	Raf1PPtase
3	0.00	4.56×10^5	BRaf, Rap1	RapGap
4	1.00×10^1	4.56×10^7	P90Rsk, PI3K, Akt	
5	1.00×10^3	4.56×10^3	Raf1	RasGap

Table 4.1: Parameter-Defining Experimental Set For Triples Design. These are the set experiments selected by the greedy search. The pallet of experiments included different doses of EGF/NGF in combination with triple knockdowns and overexpressions.

to determine a large number of parameters when applied on their own (Figure 4-2B) or if experiments were added randomly (Figure 4-2C), the performance was much worse; neither procedure could determine the full 48 parameters with up to 20 experiments, whereas the greedy experimental design procedure required only five (Figure 4-2D). Interestingly, the "random" procedure was about as effective as the "best singles" beyond the initial few experiments, which suggests that selecting "good" or "bad" experiments isn't nearly as important as choosing complementary ones For example, most of the "best singles" experiments tended to involve stimulation with EGF and NGF simultaneously (Figure 4-7), but the type of complementarity required to tease apart all the parameter directions can be achieved by more subtle variation (Table 4.1). Together these results show that selection of complementary experiments can very efficiently lead to full parameterization of complex models and that experimental design procedures are important to select an appropriately complementary set.

4.2.2 Relative Resolving Power of Different Experiment Types

There should be a cost–benefit relationship between the complexity of experimental perturbations used and the amount of information obtained. Intuitively, we expect it may take more simple experiments to obtain the same knowledge gleaned from a smaller number of more complex experiments. The above results included experiments that simultaneously modified three protein concentrations. To probe whether simpler experiments could achieve similar results, we repeated the search but limited the experiments to EGF/NGF doses only (Figure 3A,B),

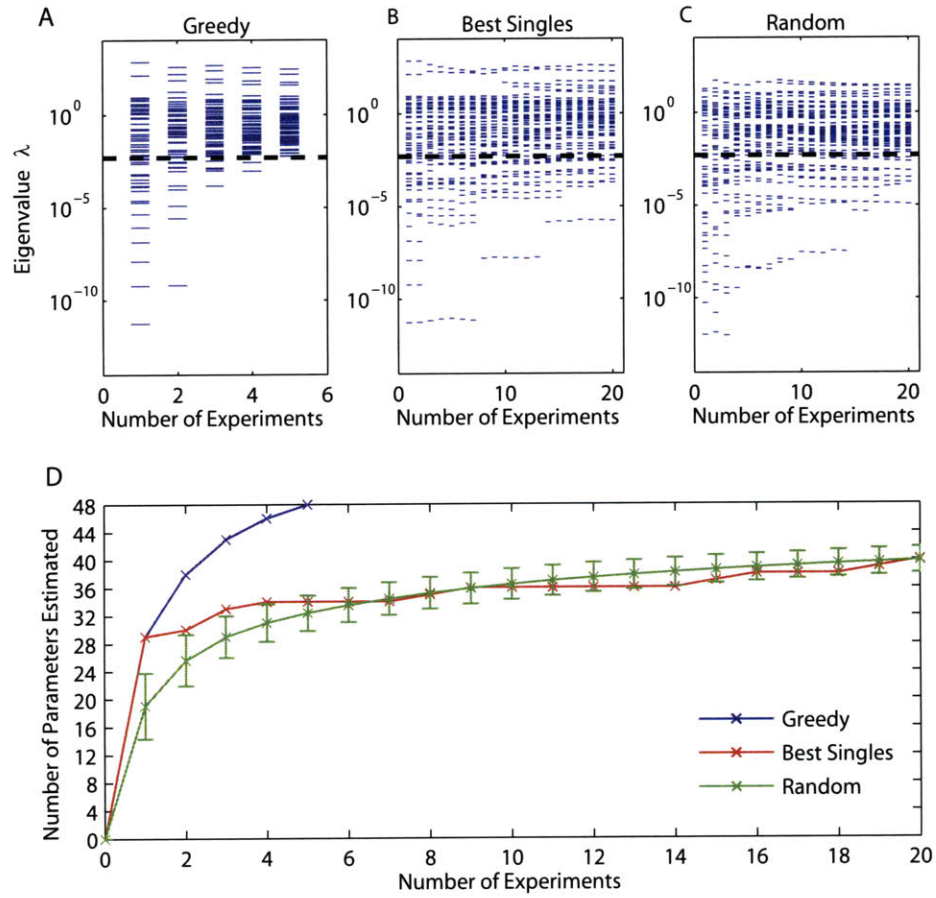


Figure 4-2: Experiment Design for EGF/NGF Model. (A) Search result for sets of experiments that maximize the number of parameters estimated. Each spectra is the eigenspectrum of the estimation problem. The goal of the design is to maximize the number of parameters with errors less than 10% (dashed line). (B) Design based on selecting the best single experiments. (C) Design based on selecting random experiments. (D) The number of parameters estimated by each search method. By the fifth experiment the greedy algorithm is able to estimate all 48 parameters to the desired accuracy. The error bars show the standard deviation for 10000 random searches.

EGF/NGF doses and single knockdowns and overexpressions (Figure 4-3C,D), or EGF/NGF doses and up to double knockdowns and overexpressions (Figure 4-3E,F). For comparison, results for the full set of experiments with up to triple expression changes are shown in Figure 4-3G,H. An experimental set with up to forty experiments was constructed for each using our greedy experimental design approach, and the maximum number of parameters determined to within 10% are shown. For each pair of figure panels, the left panel shows the location in the network of parameters established by the experimental set, with the same arrow thickness scheme as in Figure 4-7 of thicker arrows indicating more parameters defined. The eigenvector matrix, a representation of best determined to least well-determined directions in rate-constant space, is displayed in the right panel for each sub-section of Figure 4-3. The eigenvector matrix shows which individual parameters contribute to each parameter direction. All rate constants could be determined with up to triple expression changes (as shown above, with five experiments) and 47 of 48 parameters with up to double expression changes (requiring 8 experiments, with no improvement from the next 32 most complementary experiments). Most rate constants could be determined using single changes in protein expression (45 of 48, but requiring 17 experiments), but doses of EGF/NGF alone were only able to establish just over half of the rate constants (25 of 48, requiring just 2 experiments with no improvement resulting from the remaining 23 experiments in the class). Thus, more complex perturbation experiments improved model calibration through establishing a greater number of parameters and generally doing so with fewer, albeit more difficult, experiments. The tradeoff is such that in many cases the greater complexity may be more than justified by the reduced number of total experiments required. Modifications to future versions of the optimization could be biased towards reuse of genetic modifications in experiments with different dosage treatments, so that the greater effort of the former might be better leveraged.

4.2.3 Biochemical Basis for Complementarity of Experiments

Analysis of the five experiments sufficient for determining all the parameters suggests that one role for some selected experiments is to specifically adjust conditions of enzymatic re-

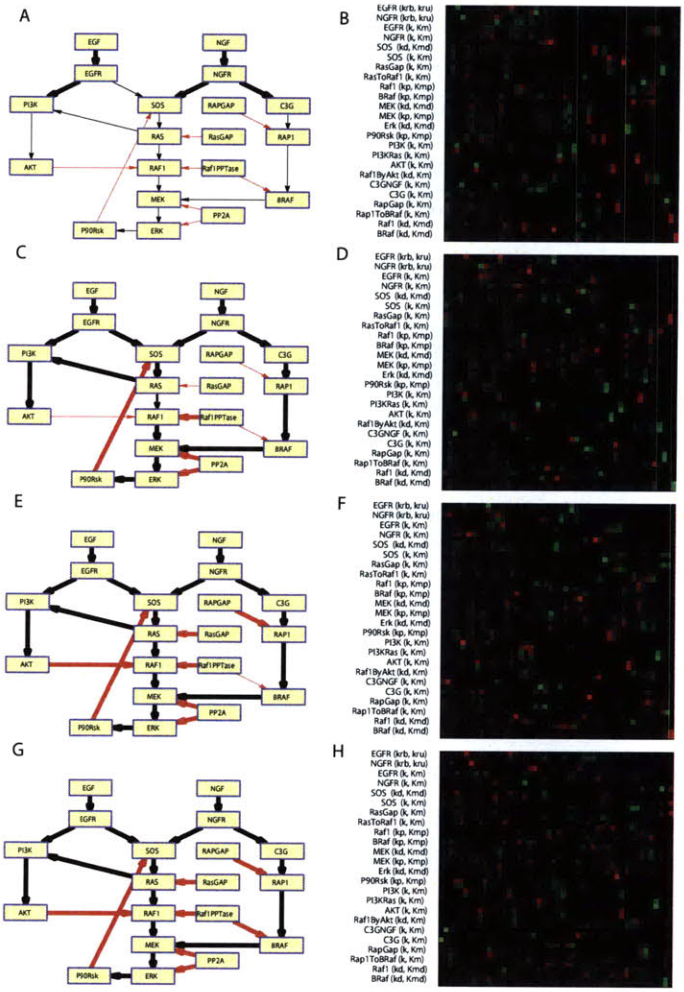


Figure 4-3: The Subset of Parameters That Can be Estimated with Different Levels of Experiment. (A-B) Doses of EGF and NGF alone. (C-D) Single knockouts/overexpressions. (E-F) Double knockouts/overexpressions. (G-H) Triple knockouts/overexpressions. The network diagrams show the individual parameter errors. Each arrow represents a reaction (black for activating, red for inhibitory). Each reaction is parameterized by two parameters. If both parameters are estimated to 10% the line is thick. If only one of the two parameters are estimated the line is medium. And if neither parameter is estimated then the line is thin. The eigenvector matrices pictures on the right show the vector perspective. The eigenvalues decrease from left to right. The green lines indicates the cutoff for 10% relative error. The yellow and red lines indicate directions with 100% and 1000% error.

actions so that k_{cat} and K_M could be independently determined. Because the calculations were done in log parameter space, the $(+k_{cat}, +K_M)$ direction corresponds to $\log(k_{cat}) + \log(K_M) = \log(k_{cat}/K_M)$. Interestingly, the $(+k_{cat}, -K_M)$ parameter direction, corresponding to $\log(k_{cat}/K_M)$, is easily estimated for many reactions here because in the wild type many enzymes operate under k_{cat}/K_M conditions with substrate concentration below K_M. Selected experimental perturbation appear to have been chosen because they drive substrate concentration sufficiently high as to move beyond k_{cat}/K_M conditions to determine k_{cat} and K_M independently (Figure 4-4)

4.2.4 More Stringent Parameterization Can be Achieved with Greater Experimental Effort

It is important to consider whether there are fundamental limits to how accurately biological models can be parameterized – essentially whether there exist model parameters that are essentially unknowable. As one step to addressing the question of knowability, we repeated the optimal experimental design calculations using up to three expression changes, but with different values of the error threshold. Our previous results were computed with the requirement that parameters be established to within 10% of their nominal values. The full set of results is shown in Figure 4-5. While five experiments were required to establish all 48 parameters to within 10% error, only four experiments were required for the less stringent 37% error. Likewise, as the stringency was increased, greater experimental effort was required to establish a given number of parameters. Together these results suggest that more stringent parameterization can generally be achieved with greater experimental effort, although experimental design procedures may be necessary to define how best to apply additional effort towards determining new knowledge. Additional experimental complexity may be necessary to dissect particularly difficult parameter combinations. Whether such effort is worthwhile may depend on the sensitivity of predictions made by model to such difficult parameter combinations.

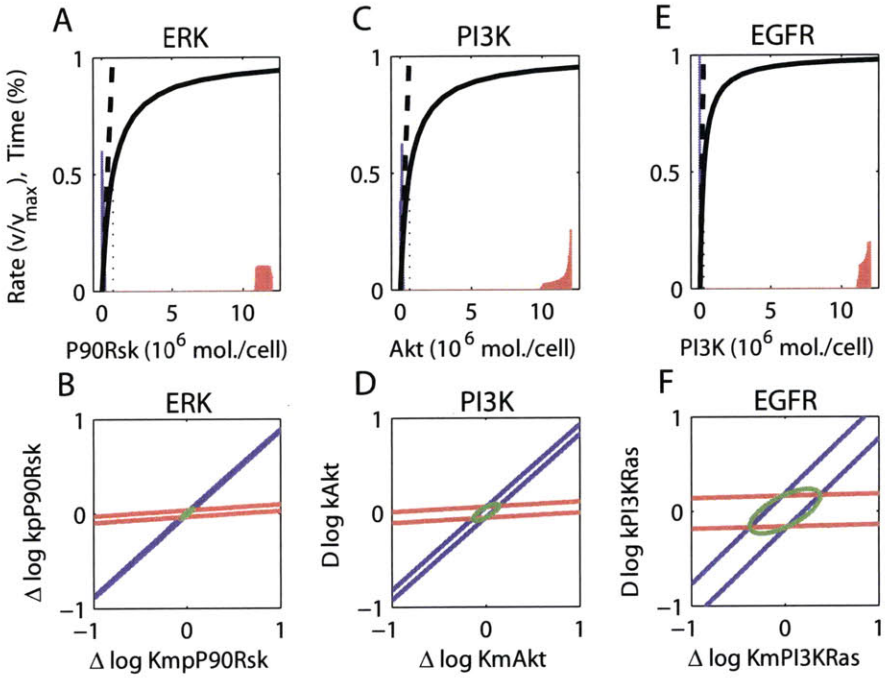


Figure 4-4: Experimental Perturbations Push Enzymes From Pperating in Pure $k_{\text{cat}}/K_{\text{M}}$ Conditions to Facilitate Estimation of k_{cat} and K_{M} Individually. Experiment #4 calls for the overexpression of the three enzymes P90Rsk, Akt, and PI3K, which are substrates for the enzymes (A,B) ERK, (C,D) PI3K, and (E,F) EGFR, respectively. The top row shows the rate of reaction for each enzyme as a function of its substrate concentration (solid black line), the linearized rate law at low substrate concentration $r \approx e_o S \frac{k_{\text{cat}}}{K_{\text{M}}}$ (dashed black line), and a histogram of concentration in experiment #1 (blue shading) and in experiment #4 (red shading). The bottom row shows the corresponding parameter uncertainty, with K_{M} on the abscissa and k_{cat} on the ordinate. Uncertainty is shown for experiment #1 alone (blue ellipse), experiment #4 alone (red ellipse), and the combination of both experiments (green ellipse). In experiment #1 only the linear region of the enzyme rate curve is explored (blue shading in top row). The linear regime specifies $k_{\text{cat}}/K_{\text{M}}$ (represented as $\log(k_{\text{cat}}) - \log(K_{\text{M}})$ in log parameter space and corresponds to the blue ellipses in the bottom row of the Figure). The blue ellipses are extended along the $y = x$ direction, indicating very little uncertainty in $\log(k_{\text{cat}}) - \log(K_{\text{M}})$ but great uncertainty in $\log(k_{\text{cat}}) + \log(K_{\text{M}})$. In experiment #4 all three substrates are overexpressed and the saturating portion of the rate curve is populated, where $r \approx e_o k_{\text{cat}}$, which specifies the $\log(k_{\text{cat}})$ but not the $\log(K_{\text{M}})$ direction, as indicated by the red ellipses being aligned nearly along the abscissa. Combining both experiments specifies both rate parameters, as indicated by the small extent of the green ellipses.

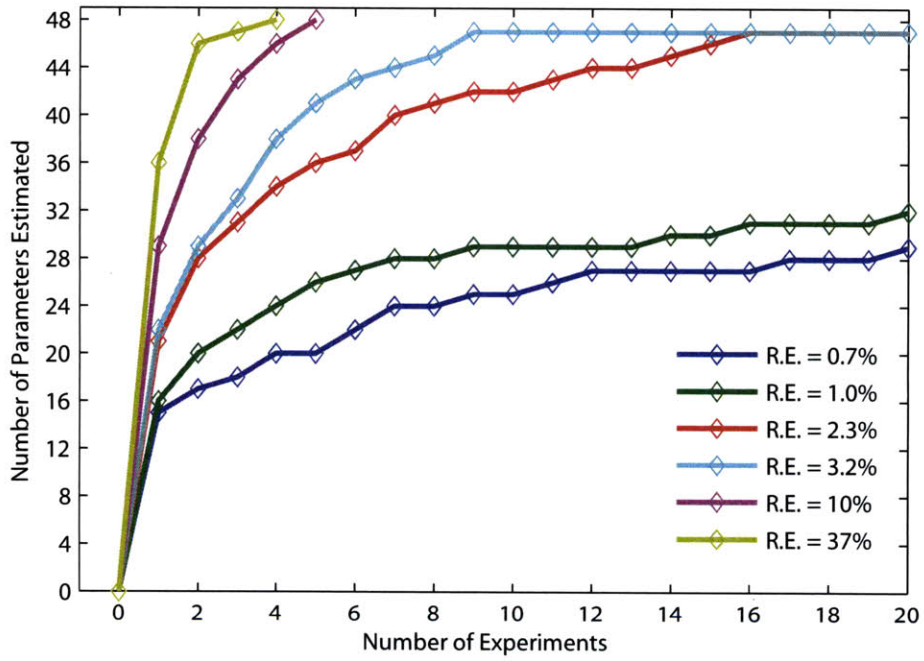


Figure 4-5: Experimental Design with Different Relative Parameter Errors. The violet line is the original design with 10% uncertainty. It is interesting to note that with an objective of 37% uncertainty (gold line) that all but two of the parameters can be estimated with two experiments, and all parameters can be estimated with four experiments.

4.3 Discussion

In this study computer simulations and experimental design methods were used to probe the relationship between experimental perturbations applied to a complex biological system and the relative certainty with which model parameters describing that system can be established. These *in silico* experiments resolve an important question in computational systems biology. They imply that uncertainty is not inherent to biological network models—rather, uncertainty can be progressively diminished through sequential addition of perturbation experiments to a cumulative data set used for model calibration. In the case studied here, a cellular network activated in response to stimulation by EGF and NGF important for cell growth and differentiation, all 48 rate constant parameters could in principle be fit to within 10% of their nominal value using concentration time courses from five multi-perturbation experiments. While the accuracy of the measurements and the desired accuracy of the parameters affect the number of required experiments and their complexity, all parameters could be estimated to very high accuracy.

An important characteristic of minimal sets of experiments that together define all model parameters is that the members are mutually complementary. That is, while they may contain some overlapping information, each member also contains information that is not provided by the others. The overall parameter uncertainty of a set of experiments is related to the intersection of the parameter uncertainty associated with each individual experiment. Thus, the intersection diagrams of Figures 1 and 4 should not be interpreted to mean that complementary experiments must contain non-overlapping large uncertainties, but rather that the directions of large uncertainty in one experiment should correspond to a direction of smaller uncertainty in at least one other experiment. This form of complementarity is non-obvious and non-trivial, but experimental design methodology can efficiently identify sets possessing this property. In the current example, running enzymes under both $k_{\mathrm{cat}}/K_{\mathrm{M}}$ and k_{cat} regimes was one important form of complementarity. Single experiments were generally incapable of spending sufficient time in each regime, and different experiments were required for each. The design of single experiments that visit both regimes could be beneficial.

The current case also exhibited a tendency for protein overexpression as opposed to knockdowns, although the generality of this result remains to be seen. Operationally, in many cases it may be possible and convenient to alter the activity of proteins with selective inhibitors while leaving the expression level constant. Small molecule inhibitors or activators also enable time-dependent perturbation, thereby providing important new degrees of freedom that may permit improved model calibration with fewer experimental manipulations [5]. However, we found certain parameter directions for which overexpressions were important for their exploration. In the work described here, we used full trajectories of all concentrations in the model in fitting; however, current experimental technology generally probes only a subset of species and time points. The experimental design framework described here can be used to determine the most productive species and time point measurements to make in order to most expeditiously calibrate model parameters.

While a variety of experimental interventions can be used to change protein expression, another possibility is to make use of natural variation. An examination of the expression levels of the proteins that correspond to the 19 proteins in the EGF/NGF signaling model as reported in the GNF SymAtlas database [166] is shown in Figure 4-8 and indicates expression ranges of 100 fold or more across tissues and cell types. Natural variation on this scale suggests that collecting data in multiple cell types may be an alternative to using genetic manipulation, but it also suggests that accurate estimation of parameters and quantification of protein expression levels may be necessary for a model to be applicable across different cell and tissue types. However, using multiple cell lines introduces additional complications, as differences in biology and un-modeled effects may dominate the results. In this case, having a well-calibrated model may make it easier to distinguish between calibration issues and real biological differences.

Many different sets of experiments can be mutually complementary and each can define all the parameters equally well. Furthermore, tradeoffs exist between the complexity of experiments available and the number of experiments required to probe fully all parameter. Multiple genetic changes were particularly effective in the EGF/NGF network studied here, with a larger number of simpler experiments performing less well overall. Here we designed sets of experiments to define the parameters of a pre-existing model, as this would allow us to most directly

address the theoretical feasibility of the approach. One result is that rather complex experiments were called for, and we were curious whether this was a necessity of the underlying biological system or an artifact of the model itself. We noticed that a few K_M values in the model were quite large, and we built an alternative model with more standard K_M values that fit the experimental as well as the original model (Figure 4-9). We applied our experimental design methodology but only permitted an experiment to consist of an EGF/NGF and 100-fold expression change of one protein; all 48 parameters could be determined using this simpler experimental set in just 15 experiments (Figure 4-9). Interestingly, in this case protein overexpression experiments dominated underexpression again. Moreover, three genetic perturbations were re-used, and in two pairs of experiments the same gene was used in both overexpression and underexpression experiments. Thus, it is likely that numerous but simple experiments can be used to quantitatively understand and describe even complex biological systems.

In our analysis we have considered the value of the Hessian only at the optimal parameter set when computing our parameter uncertainty. This corresponds to assuming that the log parameter errors are Gaussian, an assumption which may not be true for poorly determined parameter directions. However, as more complementary experiments are added, and as the parameter uncertainty is reduced, this approximation will become increasingly accurate [114].

Throughout this work, we have stressed the importance of fully determining the parameters of a model. However, as has been observed by others [78], this may not be necessary to make a particular prediction. In fact, the corollary to no single experiment determining all of the parameters is that no single prediction depends on all of the parameters. This result suggests a potential variation to our method. If a particular model was only to be used in a narrowly defined context where only certain parameters were sensitive, then the method could choose experiments to define the parameters directions that spanned the sensitive parameter directions. For example, if the substrate of an enzyme could never be expressed at a level above the K_M of any of the enzymes that modify it, you could decide to ignore the parameter errors that pointed in the $([+\log(k_{cat}), +\log(K_M)])$ direction. However, the results presented here suggest that the saving in terms of experimental effort may be minimal, as a small number of experiments was able to cover the entire parameter space, and it comes at the cost of decreasing the predictive

power of the model.

If a complementary set of experiments is used to fully parameterize a model, then there exists no perturbation to the system for which the parameterization is inadequate. This is an important statement. The identification of therapeutic approaches, drug targets, and treatment regimens essentially corresponds to identifying perturbations that produce desirable outcomes. To be meaningfully accurate, such predictions must still be adequately parameterized by the calibration experiments. The construction of a complementary set of experiments sequentially reduces all directions of parameter uncertainty. Intuitively, once all the parameters are well determined, the model will make extremely good predictions whose error can be bounded through propagation of the remaining uncertainty. (These statements are predicated on the model topology being sufficiently correct to accurately describe system dynamics and on the absence of bifurcations and other anomalies within the remaining parameter uncertainty.) On the other hand, if a model is calibrated with an incomplete set of experiments, then there exist parameter directions with large uncertainty. Any experimental prediction whose outcome is sensitive to the undetermined parameter directions, such as the effects of drug therapy, for instance, could be grossly incorrect. In fact, such an experiment might be a poor therapy but an excellent calibration tool. Our results demonstrate that it is not necessary to modulate the expression level of every protein in the network in order to determine all parameters and thus fully define the behavior of the system. Rather, only a complete set of complementary experiments need be used for model calibration; in principle, excellent predictions should follow.

4.4 Methods

4.4.1 The Model

In this work we examined a model of EGF and NGF signaling [17, 30, 31, 32, 184, 185]. The model can be found in the BioModels database (BIOMD0000000033) [106]. The model consists of 19 distinct proteins, two extracellular proteins (EGF and NGF, which act as inputs), two cell surface receptors (EGFR and NGFR), and 15 intracellular proteins. The two receptors and 11 of the cytoplasmic proteins can be either in an active or an inactive state. The remaining

four intracellular proteins are constitutively active, resulting in a total of 32 distinct chemical species, 26 chemical reactions and 48 parameters (22 Michaelis constants K_{M}; 22 catalytic rate constants k_{cat}; 2 second-order association constants, k_{on}, and 2 first-order dissociation constants k_{off}. Note that all initial concentrations are assumed to be known). Twenty-two of the reactions are implemented with Michaelis–Menten kinetics, while the remaining four reactions are the mass-action binding reactions of EGF and NGF to their respective receptors. One modification was made to the model, which was to put free EGF and NGF into an extracellular compartment having a volume 1000 times the volume of the intracellular compartment. This was chosen to better reflect a typical experiment, where the extracellular volume greatly exceeds the intracellular volume. Preliminary calculations using a variety of different volumes for the extracellular compartment demonstrated that the results did not depend strongly on the value for this volume (data not shown).

4.4.2 Objective Function

The goal of experimental design for model calibration can be expressed as selecting the experimental conditions such that the resulting parameter uncertainty ellipsoid will be prescriptively small. In this work, we focused on maximizing the number of parameters with uncertainty less than a given threshold. Equation 4.4 shows this metric as functions of the eigenvalues λ_i of $\mathbf{H}$,

$$F_N = \sum_i thresh(\lambda_i, \lambda_{thresh}) \quad \text{where} \quad thresh(\lambda_i, \lambda_{thresh}) = \begin{cases} 1 & \lambda_i > \lambda_{thresh} \\ \lambda_i/(\lambda_{\mathrm{thresh}} n_p) & \text{otherwise} \end{cases} \tag{4.4}$$

where n_p is the total number of parameter directions and $\lambda_{\mathrm{thresh}}$ is determined below. We chose to include some contribution to our objective function from parameters that did not meet the threshold to break ties between experiments with the same number of good parameter directions (thus the $\lambda_i/(\lambda_{\mathrm{thresh}} n_p)$ term instead of zero). Essentially F_N counts up the number of eigenvalues $lambda_i$ greater than a threshold. By choosing this threshold to correspond to a 10% relative parameter uncertainty, the function counts up the number of parameter directions with uncertainty less than 10%.

In this work, we assume that our data has Gaussian relative error with zero mean error and 10% standard deviation. We weighted concentration differences in Equation 4.1 by the variance in the measured value $\sigma_{s,c}(\mathbf{p}^*, t) = f \times y_{s,c}(\mathbf{p}^*, t)$, which makes our least squares estimator a maximum likelihood estimator. Here, f is the relative measurement error which we set to 0.10 (10%). Equation 4.1, describes the information from an average data point, where the average is over species, conditions, and time. Following the method of Gutenkunst et al. [78] we scale the fit metric by n_d. This corresponds to collecting n_d independent measurements of the system, where n_d is large enough so that sample mean approaches the mean. In this work, we assumed 100 times the number of species, to correspond to discreetly sampling the system at 100 time points. Note, that as the number of conditions increase we do not increase the number of data points, as this provides a useful normalization.

The variance of the parameter uncertainty can be computed by the Cramer–Rao bound [153]. The maximum likelihood estimator is asymptotically normal, unbiased, and efficient, so in the limit of a large number of data points the variance ($\sigma^2_{v_i^T \log(\mathbf{p})}$) of the log parameter uncertainty in the i$^{\text{th}}$ log parameter eigendirection (v_i) can be approximated as:

$$\sigma^2_{v_i^T \log(\mathbf{p})} = \frac{f^2}{n_d \lambda_i} \tag{4.5}$$

We define the relative parameter error to be the ratio of the upper and lower values of the parameter corresponding to the 95% or two standard deviation confidence interval [15]. Computing this involves exponentiation the bounds of log parameter error.

$$\Sigma_{v_i^T \log(\mathbf{p})} = \frac{exp(v_i^T \log(\mathbf{p}^*) + 2\sigma_{v_i^T \log(\mathbf{p})})}{exp(v_i^T \log(\mathbf{p}^*) - 2\sigma_{v_i^T \log(\mathbf{p})})} - 1 = exp(4\sigma_{v_i^T \log(\mathbf{p})}) - 1 \tag{4.6}$$

As an example $\Sigma_{v_i^T \log(\mathbf{p})} = 0.10$, means that the ratio of the maximum parameter value to the minimum parameter value is 1.1, which corresponds to relative error at the 95% confidence interval. We solved for a λ_{thresh} that corresponded to a desired parameter error by solving Equations 4.5 and 4.5. For $f = 0.10$ and $\Sigma_{v_i^T \log(\mathbf{p}^*)} = 0.10$ $\lambda_{\text{thresh}} = 0.005$. We used this value, except where noted otherwise.

4.4.3 Experimental Design

Equation 4.1 shows that the least squares objective function is a sum over conditions. Likewise, the Hessian of the least squares objective function can also be constructed as a sum of the Hessians from each individual condition,

$$H(C, \mathbf{p}) = \frac{1}{n_c} \sum_{c \in C} H_c(\mathbf{p}) \tag{4.7}$$

where C is a set of experimental conditions, c indexes the individual experimental conditions in the set, and n_c is the number of conditions in set C. Thus, we formulated the experimental design problem as the process of selecting the subset of all possible experiments such that the sum of the Hessians has the desired properties. Operationally, we simulated each individual experimental condition in the trial space and computed the Hessian. We then performed a greedy search to find the best set of n_c experiments. At each step of the algorithm, the Hessian from the previous step was added to the Hessian for each candidate experiment, and the objective function was evaluated. The trial experiment that led to the best new Hessian was added to the set. The search terminated if the goal was met or if the maximum number of experiments was reached. The choice of a greedy search does not necessarily produce the optimal subset of a given size, indicating that these results are an upper-bound on the number of experiments required to achieve a particular goal.

4.4.4 Computing Hessians

We computed the individual Hessians of this least squares cost function for each species in each condition using the MATLAB SymBio Toolbox v2.4 [169] following the method of Gutenkunst et al. [78]. The ODE system was integrated using ODE15S [156] and sensitivities were computed by complex-step finite differencing [115]. The entries in the Hessian were computed by numerical integration with the trapezoidal rule. Because we evaluated the Hessian at $\mathbf{p}^* = \mathbf{p}$, then $y_{s,c}(\mathbf{p}, t) - y_{s,c}(\mathbf{p}^*, t) = 0$ for all time points, species, and conditions; thus, the second term in the integrand of Equation 3 (containing the second derivatives of $y_{s,c}$) was also zero. For

Exp.	EGF (mol./cell)	NGF (mol./cell)	Over Expressed	Knocked Down
1	1.00×10^7	4.56×10^7		PP2A
2	1.00×10^1	4.56×10^3		RasGap
3	1.00×10^7	4.56×10^3	P90Rsk	
4	0.00	4.56×10^5		RapGap
5	1.00×10^7	4.56×10^5	Sos	
6	1.00×10^7	4.56×10^5	Akt	
7	0.00	4.56×10^3	PI3K	
8	1.00×10^3	4.56×10^1	BRaf	
9	1.00×10^7	0.00	Mek	
10	1.00×10^7	4.56×10^5	Raf1	
11	1.00×10^7	0.00	PI3K	
12	1.00×10^7	4.56×10^1	Erk	
13	1.00×10^7	4.56×10^1	PP2A	
14	1.00×10^7	4.56×10^3	Mek	
15	1.00×10^7	4.56×10^7	Sos	

Table 4.2: Parameter-Defining Experimental Set For Singles Design.

the purpose of this study we dropped this term. Note that even with imperfect data it may be possible to approximate the Hessian using only first derivatives if the fit to the data is sufficiently good that the second derivative term is negligible. If this is not the case then the full Hessian equation should be used, and the rest of the analysis remains unchanged. The computations were performed in parallel on 128 processors using approximately 800 cpu hours in total.

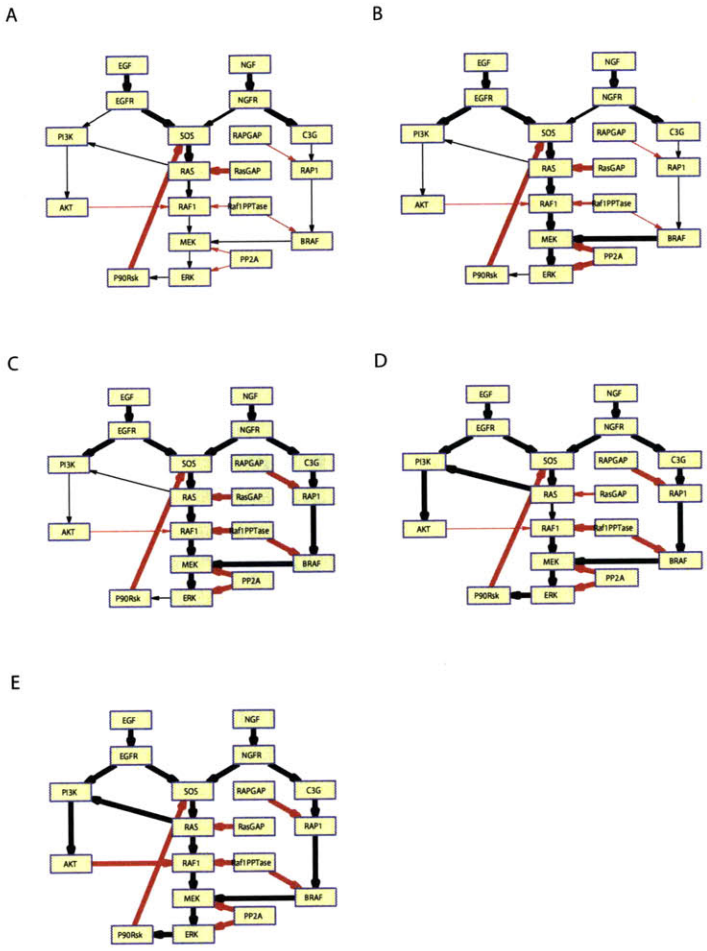


Figure 4-6: Parameter Elucidation Determined by Greedy Search. Thick lines indicate both parameters associated with the reaction were determined to within 10% by the bounding box metric. Medium lines indicate that one of the two parameters associated with the reaction were determined to within 10%. Thin line indicate that neither were determined to within 10%. (A) Experiment 1 only, (B) Experiments 1 and 2. (C) Experiments 1-3 (D) Experiments 1–4, (E) Experiments 1–5. Overall the first parameters discovered were in the core SOS to ERK pathway followed by the C3G to BRAF then the PI3K, AKT pathway.

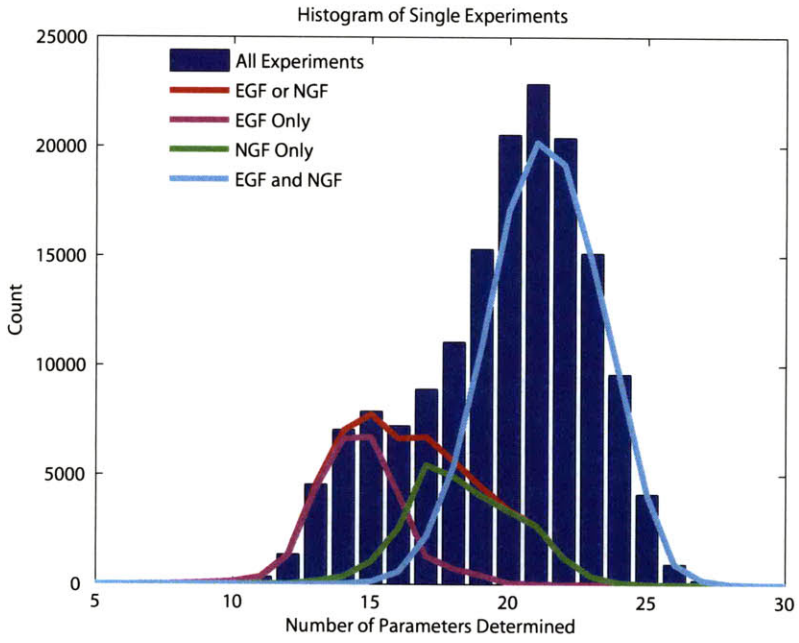


Figure 4-7: Histogram of the Number of Parameters Determined to 10% Relative Error for Each Single Experiment. The blue bars shows the results for all experiments; the cyan line is the subset of experiments that used an EGF and a NGF stimulus; the red line are the subset that used only one stimulus (EGF or NGF); the purple lines is for EGF alone; and green line is for NGF alone. In general, the double stimuli performed better than the singles, but not twice as good, and there is a long tail to the single subset.

	Parameter Value	
Parameter Name	Original	Modified
kEGF	694.731	41.0944
KmEGF	6.09×10^6	3.60×10^5
kdSos	1611.97	647.019
KmdSos	8.96896×10^5	3.60×10^5
kRasGap	1509.36	77.6942
KmRasGap	1.43×10^6	73733.3
kpRaf1	185.759	70.122
KmpRaf1	4.77×10^6	1.80×10^6
kpMekCytoplasmic	9.85367	7.652
KmpMekCytoplasmic	1.01×10^6	782264
kdErk	8.8912	4.14072
KmdErk	3.50×10^6	1.63×10^6
kpP90Rsk	0.0213697	0.0100758
KmpP90Rsk	7.63523×10^5	3.60×10^5
kAkt	0.0566279	0.0311737
KmAkt	6.53951×10^5	3.60×10^5
kdRaf1ByAkt	15.1212	1.99935
KmRaf1ByAkt	1.19355×10^5	15781.3
kRapGap	27.265	4.04006
KmRapGap	2.9599×10^5	43859.1
kRap1ToBRaf	2.20995	0.775829
KmRap1ToBRaf	1.03×10^6	3.60×10^5
kdBRaf	441.287	0.0832448
KmdBRaf	1.09×10^7	2052.32

Table 4.3: Modified Parameters at Alternative Parameterization Parameter Value. Catalytic rates start with a 'k' and are in units of molecules (cell–volume)$^{-1}$ s^{-1}. Michaelis constants start with a 'Km' and have units molecules (cell–volume)$^-$1.

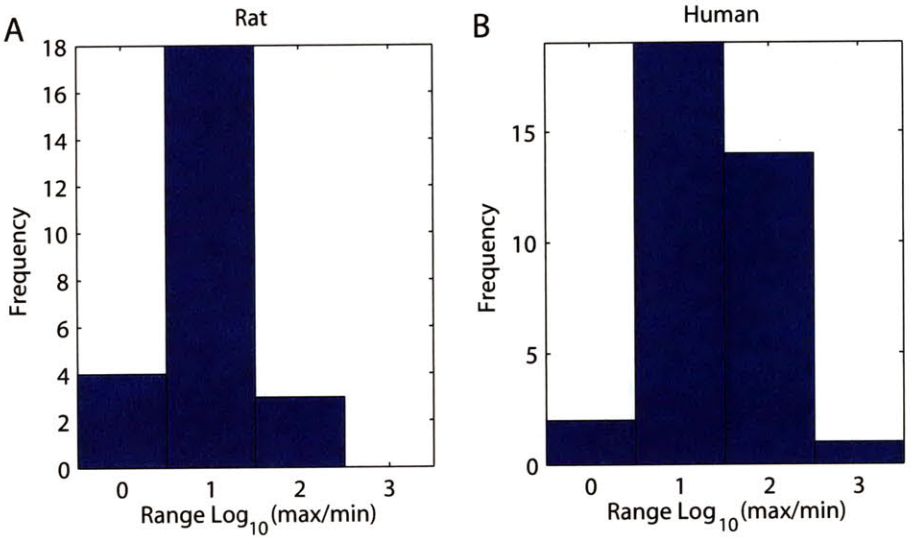


Figure 4-8: Histogram of Expression Levels for Genes that Code for the Proteins in the Model. (A) Results from different cell types and tissues in Rats. (B) Results for various cell types and tissues in Humans. The expression levels of almost all of the genes vary by 10-fold or more with about half varying 100-fold.

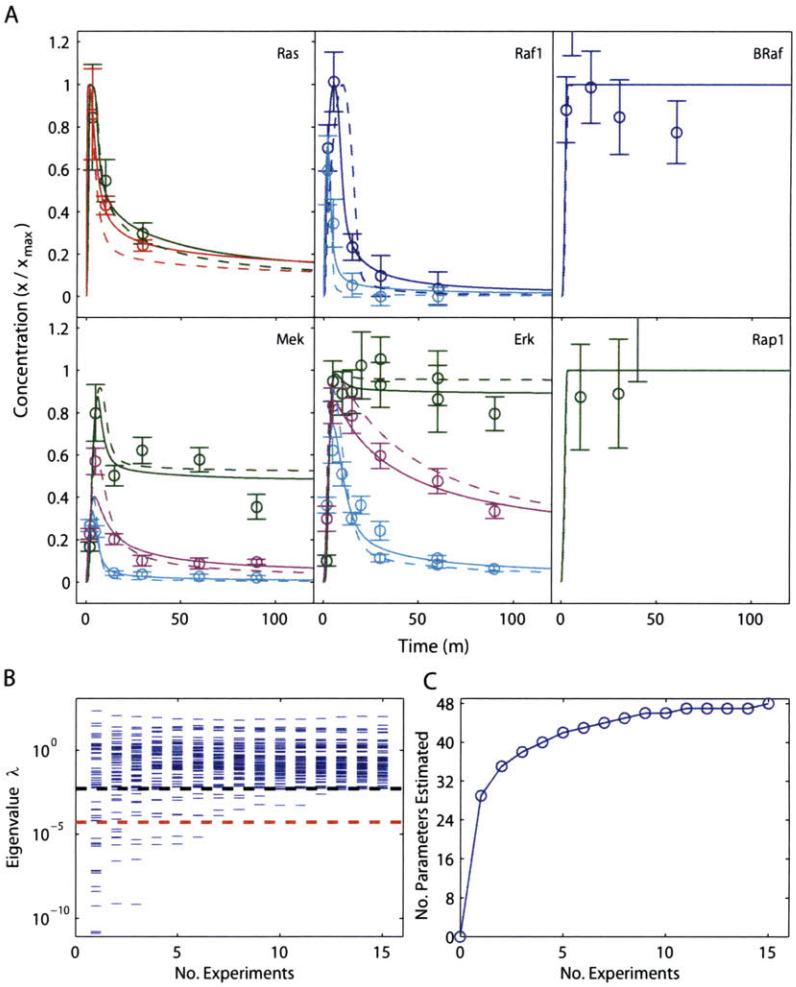


Figure 4-9: Design with Single Genetic Perturbations at Another Feasible Parameterization. The parameters of the model were changed such that K_M for each enzyme could not be greater than three its maximum substrate concentration. (A) The results of simulating the model at the original parameterization (solid lines) the modified parameterization (dashed lines), and experimental data (circles with error bars). The line colors indicate the experimental condition: 100 ng/mL NGF (blue), 50 ng/mL NGF (green), 30 ng/mL EGF (red), 100 ng/mL EGF (cyan), 100 ng/mL EGF with 50x EGF overexpression (magenta). The experiment design methodology was applied with EGF/NGF doses and single gene changes. (B) Eigenspectrum of experiments. (C) Number of parameters determined by each step of the search. With the modified parameterization all 48 of the parameters could be determined to 10% R.E. (black dashed line) with 15 experiments, and to within 100% R.E with seven experiments (red dashed line).

Chapter 5

Formulation of Mass Action Models of Biological Systems Using Kronecker Products[1]

5.1 Abstract

Mechanism-based chemical kinetic models are increasingly used to describe biological signaling. Such models serve to encapsulate current understanding of pathways and to enable insight into complex biological processes. Currently, there is no standard formulation for these models, making it difficult implement general-purpose analytical and computational methods. Here, we propose a general and concise formulation of mass action kinetics based on sparse matrices and Kronecker products. We show that this formulation allows any mass action model and its partial derivatives to be represented by simple matrix equations, allowing for the straightforward application of numerical techniques. Based on this result, we implemented three standard analytical methods: dose response analysis, model simulation and calibration. In each of these, we exploited our formulation to derive an efficient implementation in a high-level mathematics

[1]This chapter is adapted from a manuscript currently in preparation: Apgar JF*, Toettcher JE*, Tidor B, and White J. Formulation of Mass Action Models of Biological Systems Using Kronecker Products. * These authors contributed equally to this work.

package. For example, sensitivities are computed by an adjoint method. We applied these methods to a series of models which demonstrated that they scale to models of thousands of equations on standard desktop hardware.

5.2 Introduction

One goal of systems biology is to develop detailed models of complex biological systems that quantitatively capture known mechanisms and behaviors, and also make useful predictions. Such models serve as a basis for understanding, for the design of experiments, and for the development of clinical intervention. In support of this goal, there has been a strong push to build mechanistically correct kinetic models, often based on systems of ordinary differential equations (ODEs), that are capable of recapitulating the dynamic behavior of a signaling network. These models hold the promise of connecting biological and medical research to a class of computational analysis and design tools that could revolutionize how we understand biological processes and develop clinical therapies [101, 97].

Data formats such as Systems Biology Markup Language (SBML) [62, 88] and Cell markup language (CellML) [112] have standardized the representation of biological models, allowing them to be shared between laboratories and encouraging the development of model repositories. This has greatly increased the number of publicly available models [129]. However, unlike the models themselves, computational methods have remained difficult to share. For the most part they rely on one–off, and often proprietary codes. As a result the methods have remained siloed within individual software packages and are difficult to combine, customize, or apply in new contexts. The problem is exacerbated by the fact that CellML and SBML allow for arbitrary mathematical expressions for the rate laws. Designing general codes for the analysis of models that may contain arbitrary mathematical structure leads to high computational overhead, or spotty support.

However, the fundamental chemistry, as described by the mass action kinetics of elementary reactions, is more limited in mathematical structure. Working within this reduced instruction set allows for efficient and straightforward implementation of numerical methods. Furthermore,

this simplified basis allows for a standardized model representation, which can easily be implementation in a high-level language. Together this lends itself to the development of efficient numerical methods that can be ported between projects.

We present here a standardized formulation of mass action chemical kinetics as a simple matrix equation (Equation 5.6). This formulation has the advantage that the matrices are constant and depend linearly on the rate parameters, and that the quadratic terms are expressed as Kronecker products. We show that this formulation allows simple algebraic manipulation of the models and can serve as the basis for efficient general-purpose codes. In particular, we formulate numerical integration, steady state finding, sensitivity analysis, and model calibration. Despite being implemented in a high-level mathematics package [168], we were able to obtain good performance on models with thousands of equations.

5.3 Background

5.3.1 Chemical Kinetics and the Law of Mass Action

In chemical kinetics, the law of mass action describes the mathematical relationship between the concentration of reactant species and the rate of reaction [178]. In this work, we consider mass action kinetic models consisting of zeroth-, first-, and second-order reactions described by ordinary differential equations. In the equations below, k signifies a rate constant; A, B, and C represent species or concentrations of species, depending on the context; and $\emptyset$ is the empty set or nothing.

Zeroth-order reaction:

$$\emptyset \xrightarrow{r} A \qquad \frac{dA}{dt} = r \tag{5.1}$$

First-order reaction:

$$A \xrightarrow{r} B \qquad -\frac{dA}{dt} = \frac{dB}{dt} = r\ A \tag{5.2}$$

Second-order reaction:

$$A + B \xrightarrow{r} C \qquad -\frac{dA}{dt} = -\frac{dB}{dt} = \frac{dC}{dt} = r\ A\ B \tag{5.3}$$

5.3.2 Biological Models

Biochemical models are often described as networks of chemical reactions where the time derivative of any species is given by sum over all reactions that species participates in. Mathematically these networks can be represented as systems of ordinary differential equations.

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}\left(\mathbf{x}, \mathbf{u}\left(t\right), \mathbf{r}\right) \tag{5.4}$$

$$\mathbf{x}\left(0\right) = \mathbf{x}_0 \tag{5.5}$$

If the model consists solely of reactions of the form of Equations 5.1-5.3 then $\mathbf{f}$ can be represented compactly as shown in Equation 5.6.

$$\mathbf{f} = \mathbf{A}_1\mathbf{x} + \mathbf{A}_2(\mathbf{x} \otimes \mathbf{x}) + \mathbf{B}_1\mathbf{u} + \mathbf{B}_2(\mathbf{x} \otimes \mathbf{u}) + \mathbf{k} \tag{5.6}$$

The state vector $\mathbf{x}$ describes the chemical species concentrations that are free to evolve in time according to the kinetics of the system. The input vector $\mathbf{u}\left(t\right)$ represents the chemical species concentrations controlled by the experimenter. Matrices $\mathbf{A}_1\left(\mathbf{r}\right)$ and $\mathbf{B}_1\left(\mathbf{r}\right)$ represent first-order reactions, matrices$\mathbf{A}_2\left(\mathbf{r}\right)$ and $\mathbf{B}_2\left(\mathbf{r}\right)$ represent second-order reactions, and $\mathbf{c}\left(\mathbf{r}\right)$ represents constitutive (zeroth-order) reactions. The symbol $\otimes$ denotes the Kronecker product (also known as the matrix direct product) [75]. For vectors, this operator generates a vector of all quadratic products (Equation 5.7).

$$\mathbf{x} \otimes \mathbf{u} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \otimes \begin{bmatrix} u_1 \\ \vdots \\ u_m \end{bmatrix} = \begin{bmatrix} x_1u_1 \\ \vdots \\ x_1u_m \\ x_2u_1 \\ \vdots \\ x_nu_m \end{bmatrix} \tag{5.7}$$

The output of the model $\mathbf{y}$ is a linear combination of the state variables represented by the

matrix $\mathbf{C}$.

$$\mathbf{y} = \mathbf{C}\,\mathbf{x} \tag{5.8}$$

Finally, such a model can be fully parameterized by $\mathbf{p}$, the augmented set of rate constants and initial species concentrations (Equation 5.5).

$$\mathbf{p} = \begin{bmatrix} \mathbf{r} \\ \mathbf{x}_0 \end{bmatrix} \tag{5.9}$$

We define number of parameters, species, inputs and outputs as $n_{\mathbf{x}}$, $n_{\mathbf{p}}$, $n_{\mathbf{u}}$ and $n_{\mathbf{y}}$, respectively.

5.3.3 Partial derivatives of the model

Many analysis methods applied to such models require the computation of various partial derivatives of the system of Equations 5.4. Using properties of the Kronecker product, and the fact that the $\mathbf{A}_i$ and $\mathbf{B}_i$ matrices and $\mathbf{k}$ vector are linearly dependent on rate constants, the first- and second-order partial derivatives can be written as shown in Equations 5.10–5.14.

$$\mathbf{f}_{\mathbf{x}} = \mathbf{A}_1 + \mathbf{A}_2(\mathbf{I} \otimes \mathbf{x} + \mathbf{x} \otimes \mathbf{I}) + \mathbf{B}_2(\mathbf{I} \otimes \mathbf{u}) \tag{5.10}$$

$$\mathbf{f}_{\mathbf{p}} = \mathbf{A}_{1,\mathbf{p}}\mathbf{x} + \mathbf{A}_{2,\mathbf{p}}(\mathbf{x} \otimes \mathbf{x}) + \mathbf{B}_{1,\mathbf{p}}\mathbf{u} + \mathbf{B}_{2,\mathbf{p}}(\mathbf{x} \otimes \mathbf{u}) + \mathbf{k}_{\mathbf{p}} \tag{5.11}$$

$$\mathbf{f}_{\mathbf{xx}} = 2\mathbf{A}_2 \tag{5.12}$$

$$\mathbf{f}_{\mathbf{pp}} = \mathbf{0} \tag{5.13}$$

$$\mathbf{f}_{\mathbf{xp}} = \mathbf{A}_{1,\mathbf{p}} + \mathbf{A}_{2,\mathbf{p}}(\mathbf{I} \otimes \mathbf{x} + \mathbf{x} \otimes \mathbf{I}) + \mathbf{B}_{2,\mathbf{p}}(\mathbf{I} \otimes \mathbf{u}) \tag{5.14}$$

In Equations 5.11 and 5.14, we define the augmented matrices $\mathbf{A}_{i,\mathbf{p}}$ and $\mathbf{B}_{i,\mathbf{p}}$ as the stacked matrices shown in Equation 5.15, and reshape the right-hand sides to the appropriate matrix sizes.

$$\mathbf{M}_{i\mathbf{p}} = \begin{bmatrix} \frac{\partial \mathbf{M}_i}{dp_1} \\ \vdots \\ \frac{\partial \mathbf{M}_i}{dp_N} \end{bmatrix} \tag{5.15}$$

Care must be taken to reshape the second derivatives of Equations 5.12–5.14 to consist of the proper order of partial derivatives; following [136] for all such second derivatives, we use the convention of Equation 5.15. Similarly, this matrix can be reshaped to give the $(n_{\mathbf{p}}n_{\mathbf{x}}) \times n_{\mathbf{x}}$ matrix $\mathbf{f}_{\mathbf{px}}$.

The advantage of our formulation is that the model and its partial derivatives can be succinctly represented by a small number of matrix equations. Such a simple, abstract representation obviates the need for complex tools such as automatic code differentiation, and allows the straightforward application of numerical techniques.

5.4 Analysis Applications

5.4.1 Dose response analysis

Biological system models are commonly analyzed by computing a dose response curve, defined as the steady state output of the system as a function of input concentration. Solving for steady states can be accomplished by solving for the concentration $\mathbf{x}$ that makes the right hand side of Equation 5.6 equal to zero. This problem can be solved by Newton's method.

$$\mathbf{f}_{\mathbf{x}}(\mathbf{x}^{(i)}, \mathbf{u}, \mathbf{p})\Delta\mathbf{x} = -\mathbf{f}(\mathbf{x}^{(i)}, \mathbf{u}, \mathbf{p}) \tag{5.16}$$

$$\mathbf{x}^{(i+1)} = \mathbf{x}^{(i)} + \Delta\mathbf{x} \tag{5.17}$$

For many biological networks $\mathbf{f}_{\mathbf{x}}$ (also called the Jacobian) is not full rank: linearly de-

pendent rows correspond to mass conservation relationships within the network. As a result, Equation 5.16 cannot be solved directly. One refinement to this method is to compute the singular value decomposition of $\mathbf{f_x}$ and solve over its non-null subspace. An alternative solution is to simultaneously minimize the flux with mass-constraint violation $\mathbf{Mx} - \mathbf{Mx_0}$ [58, 150] where the matrix Mx is a constant for all trajectories of the system .

$$\min_{\mathbf{x}\in\mathbf{X}} \left\{ \mathbf{f}^T\mathbf{f} + (\mathbf{Mx} - \mathbf{Mx}_0)^T(\mathbf{Mx} - \mathbf{Mx}_0) \right\} \tag{5.18}$$

The minimization problem of Equation 5.18 can be solved by gradient descent methods. This requiring the gradient $\mathbf{G}$ given by Equation 5.19

$$\mathbf{G} = 2\mathbf{f}^T\mathbf{f_x} + 2\left(\mathbf{Mx} - \mathbf{Mx}_0\right)^T\mathbf{M} \tag{5.19}$$

Timing results for both methods are presented in Section 5.6.

5.4.2 Simulation

Another analysis technique often applied to biological models is the time integration of the model equations, often referred to simply as simulation or "running the model". This may be the autonomous system response, or the response to a time varying stimulus. We have implemented an implicit integration method using a commercial stiff ODE solver `ode15s` [156]. Stiff solvers rely on the the ability to efficiently compute the time derivative of x and the Jacobian of the system (f_x) [35, 6]. Both of of these quantities can be computed efficiently for our formulation (Equation 5.10).

5.4.3 Sensitivity Analysis

Often, the modeler is interested in computing the full state sensitivities$\mathbf{x_p}(t)$, the sensitivity of the smaller set of model outputs $\mathbf{y}$, or the sensitivity of an auxiliary function $\Psi(\mathbf{p})$ (e.g. a scalar cost function for optimization). These quantities can be computed by solving the system

of Equations 5.20, and 5.21, consisting of a system of ODEs of size $n_{\mathbf{p}} \times n_{\mathbf{x}}$.

$$\frac{d}{dt} x_{\mathbf{p}} = f_{\mathbf{x}} x_{\mathbf{p}} + f_{\mathbf{p}} \tag{5.20}$$

$$x_{\mathbf{p}}(0) = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n_{\mathbf{x}}} \end{bmatrix} \tag{5.21}$$

In cases where the full concentration sensitivities are not required, an adjoint formation of the sensitivity equations can greatly speed up the calculation. For ODE systems of the form of Equation 5.6 the adjoint sensitivities are given by Equations 5.22, 5.23, and 5.24 [36].

$$\frac{d\lambda^*}{dt} = -\lambda^* \mathbf{f}_{\mathbf{x}} - \mathbf{g}_{\mathbf{x}} \tag{5.22}$$

$$\lambda^*(T) = \mathbf{0} \tag{5.23}$$

$$\frac{d\Psi(\mathbf{p})}{d\mathbf{p}} = \int_0^T \left(\lambda^* \mathbf{f}_{\mathbf{p}} + \mathbf{g}_{\mathbf{p}}\right) dt + \lambda^*(0)\, \mathbf{x}_{\mathbf{p}}(0) \tag{5.24}$$

Whereas the forward sensitivity method involved integrating $n_{\mathbf{p}} \times n_{\mathbf{x}}$ coupled ODEs, the adjoint method requires solving $n_{\mathbf{x}}$ coupled ODEs followed by $n_{\mathbf{p}}$ integrals for a total of $n_{\mathbf{p}} + n_{\mathbf{x}}$ equations.

5.4.4 Hessian Vector Products

Gradient-based minimization methods require the evaluation of the gradient of the objective as well as directional second-order derivatives (computed as the product of the Hessian and a direction vector $\mathbf{v}$ (Equation 5.16). An adjoint formulation[136, 137] also allows the efficient computation of these Hessian-vector products.

$$\dot{\mathbf{s}} = \mathbf{f}_{\mathbf{x}} \mathbf{s} - \mathbf{f}_{\mathbf{p}} \tag{5.25}$$

$$\mathbf{s}(0) = \mathbf{x}_{\mathbf{p}}(0)\, \mathbf{v} \tag{5.26}$$

$$\dot{\mu} = -\mathbf{g_{xx}s} - \mathbf{g_{xp}v} - \mathbf{f_x}\mu - (\lambda \otimes \mathbf{I}_{n_x})(\mathbf{f_{xp}u} + \mathbf{f_{xx}s}) \tag{5.27}$$

$$\mu(T) = \mathbf{0} \tag{5.28}$$

$$\frac{d^2\Psi}{d\mathbf{p}^2}\mathbf{v} = \int_0^T \mathbf{g_{px}s} + \mathbf{g_{pp}v} + \mathbf{f_p}\mu + (\lambda \otimes \mathbf{I}_{n_p})(\mathbf{f_{pp}u} + \mathbf{f_{px}s})\, dt \tag{5.29}$$

5.4.5 Model calibration

One important problem in model construction and validation is model calibration. In this, model parameters and initial conditions are varied so as to cause the model to match experimental observations. Typically the level of agreement between the model and the experimental observations is expressed as a scalar metric, such as the sum of the squares error of the observed outputs versus the predicted outputs. These observations may be continuous (Equation 5.30) or discrete (Equation 5.31).

$$\Psi(\mathbf{p}) = \int_0^T \|\mathbf{y}(\mathbf{x}(\mathbf{p},t)) - \mathbf{y}_{\text{data}}(t)\|_2^2\, dt \tag{5.30}$$

$$\Psi(\mathbf{p}) = \sum_i \|\mathbf{y}(\mathbf{x}(\mathbf{p},t_i)) - \mathbf{y}_{\text{data}}(t_i)\|_2^2 \tag{5.31}$$

In either case the calibration problem can be phrased as a nonlinear optimization problem (Equation 5.32).

$$\min_{lb_i<p_i<ub_i} \Psi(\mathbf{p}) \tag{5.32}$$

We have implemented model calibration using MATLAB's `fmincon`, a constrained nonlinear optimizer. We configured `fmincon` to use the active set algorithm [70] which is a Quasi-Newton method. Gradients of the objective function are computed using adjoint methods as described in Section 5.4.3.

5.5 Implementation Details

We implemented the numerical procedures described in Section 5.4 and applied them to three classes of biochemical models: a discretized PDE of antibody penetration into a tumor, the enzymatic cascade (EC) consisting of an interacting chain of the modules studied in [73], and a canonical MAP kinase model [87].

Our PDE model consists of a simple binding reaction in which a mobile antibody binds to immobilized tumor antigen. In this model, the unbound antibody is free to diffuse through the tumor. We model this process as a discretized PDE, noting that diffusion can be represented within the mass action framework as a first-order reaction. We vary the number of compartments to generate models of different scale.

The enzymatic cascade model is comprised of a chain of enzymatic push-pull loops. For each loop there is an activating enzyme which activates a substrate and a deactivating enzyme that deactivates the substrate. The active substrate is the activating enzyme for the next loop in the cascade. As with the PDE example, we generate models of different scales by changing the number of loops in the cascade.

5.5.1 Sparse Matrix Implementation

The matrices specifying mass action models can be very large (for example, the $\mathbf{A}_2$ matrix size scales as $n_{\mathbf{x}}^3$). However, a typical biological model does not densely populate these matrices. Physically, this derives from the fact that each chemical species does not directly react with all other species. The total number of non-zero entries in the $\mathbf{A}_1$, $\mathbf{A}_2$, $\mathbf{B}_1$, and $\mathbf{B}_2$ matrices scales with $\mathcal{O}(m \cdot n_{\mathbf{x}})$, where m is the average number of reactions participated in by each species. For protein-protein interaction networks, a typical value of m might be 3.5-7.0 [79], a range consistent with the models used in this study (Table 5.1). Thus, we choose a sparse implementation for all matrices.

Model	Reactions Per Species
MAPK [87]	3.1
PDE	3.0
EC [73]	2.8

Table 5.1: Average number of reactions per species for the models implemented in this work.

Model	No. Modules	Not Pre-computed	Pre-computed
PDE	500	24.307 s	4.862 s
EC	100	13.511 s	2.998 s

Table 5.2: Simulation Results With or Without the Pre-computed Kronecker Products. The not pre-computed column shows the timing results for simulations that use sparse matrices. The pre-computed column shows timing results that used sparse matrices and pre-computation of the terms in $x \otimes x$ and $x \otimes I$ that were needed. The result is a 5-fold improvement for the PDE model and 4.5-fold improvement for the enzyme cascade.

5.5.2 Sparse Kronecker Product

In addition to overall sparsity of the $\mathbf{A}_2$ matrix, most of the columns of $\mathbf{A}_2$ are all zero. To further improve performance, we pre-compute the terms in the $x \otimes x$ vector and the $x \otimes I$ and $I \otimes x$ matrices that will be multiplied by non-zero columns of $\mathbf{A}_2$. We only evaluate these non-zero products. The effect of this pre-computation on simulation times is shown in Table 5.2.

Because scalar multiplication is commutative, the Kronecker product $x \otimes x$ double-counts products by including both x_i, x_j and x_j, x_i terms. We chose to implement second-order reactions by dividing the flux between each of the two terms to preserve the symmetry of this matrix. We do not find that this choice significantly affects our computational capabilities; however, such a refinement could be easily implemented.

5.6 Results

To validate an implementation of the modeling framework and methods described above, we performed dose response analysis, simulation, and calibration of models of varying size and complexity. All results were generated on an Intel Core 2 Duo 1GHz desktop with 2 GB RAM running Windows XP.

Model	SVD + Newton's Method	Constrained Optimization
EC, 100N	0.13 s	7.32 s
MAPK	N/A	19.6 s

Table 5.3: Dose Response Curve Results for Two Biochemical Models. Inputs were swept over four orders of magnitude for each model. N/A indicates that the method did not converge.

Number of species	PDE	EC
N	0.0302 s	0.0560 s
5N	0.0332 s	0.1016 s
10N	0.0343 s	0.1185 s
50N	0.0834 s	0.7823 s
100N	0.2502 s	2.969 s
500N	4.8688 s	80.95 s
1000N	19.231 s	365.4 s

Table 5.4: Simulation Run Times in Seconds for Models of Varying Size. The number of equations per model scales with the number of sub-models (e.g. PDE compartments or EC cycles) with the number of equations per sub-model N.

5.6.1 Newton's Method and Dose Response Analysis

We ran both dose response tools on the MAPK model and 100-step EC model. The results are summarized in Table 5.3. In general, we find that the SVD method is faster but that it fails to converge on some problems where the constrained optimization can find a solution.

5.6.2 Simulation

To demonstrate that our implementation was capable of integrating large models, we implemented a series of models of increasing size. The results of this study are shown in Table 5.4. We find that simulation time grows less than quadratically for models of up to thousands of species. This can be seen in Figure 5-1 as a slope of less than two on the log–log plot. It should be noted that biochemical models comprising thousands of equations are rarely found in the current literature. Our results validate that a sparse matrix implementation was crucial for handling models of this size, as the dense matrices would not have fit into memory on the machine used.

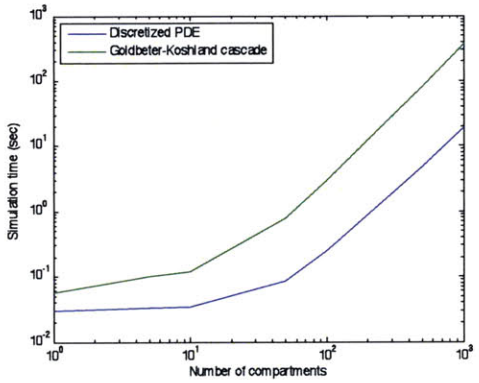


Figure 5-1: Simulation Run Times in Seconds for Models of Varying Size.

5.6.3 Model Calibration

We generated data from a nominal parameter set. These data comprised the full time-dependent trajectories for the model outputs which were a subset of the full state trajectories. The data was generated using a nominal parameter set. We then calibrated the model to the data beginning with a randomized initial guess. Optimization was performed using the integrated squared error objective function (Equation 5.30) and the active set algorithm of `fmincond`. Figure 5-2 shows the objective function at each iteration of the algorithm for three models. We find that the method converges super-linearly. While the details of convergence do depend on the initial guess, we find that the method converged to a good solution for nearly all random parameterizations tested (data not shown).

5.7 Conclusions

In this work, we have developed a concise mathematical formulation of mass action kinetics based on sparse matrices and Kronecker products. We have shown that this formulation leads to a straightforward implementation of many numerical methods. In particular, we have demonstrated that dose response analysis, simulation and model calibration can be performed efficiently within this framework for a variety of models of up to thousands of equations. This is surprising given that all of these methods are implemented in MATLAB, high-level mathematics

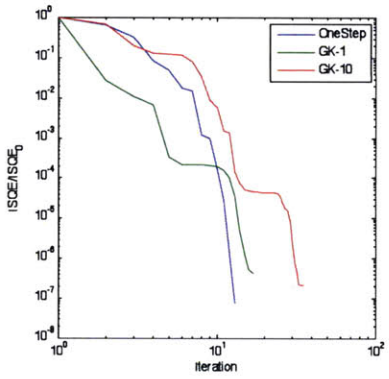


Figure 5-2: Convergence Curves for Model Calibration Using Three Models. For all three models the integrated squared error ($ISQE$) falls rapidly. The results are normalized by the $ISQE_0$, the value of the metric at the initial guess. By the 13th iteration all three models have improved by a factor of 10,000.

package.

In this work, we have focused on a set of core numerical methods. These general methods have application in almost any systems biology project. In addition these methods can also serve as the underpinning of more specialized methods. One example is that parameter optimization methods described here form the basis of the dynamic optimization controller used for model selection (Chapter 2). Another example, is sensitivity analysis code is the basis for experiment design for parameter estimation (Chapter 4). Finally, the optimization methods presented here can be applied to other features of biological systems aside from just the concentrations of species at particular times. For example, this code was adapted to allow for optimization based on the timing of particular events in the cell cycle [170]. Taken together, these results demonstrate the utility, and flexibility of the methods presented here.

Chapter 6

Converting Aggregated Rate Law Models to Mass Action

6.1 Introduction

Dynamic cell signaling and metabolic network models are predominantly implemented with ARLs. These laws simplify complex enzyme mechanisms by aggregating systems of elementary reactions into a single rate law. For example, the stepwise reaction [119] by which an enzyme binds a substrate, modifies that substrate, and then releases it is represented as a single aggregated reaction with a Michaelis–Menten rate law [123, 26]. One class of simplification is based on separation of time scale arguments such as rapid equilibrium [154, page 18], or steady state [154, page 505]. Another class is based on approximations that hold over certain concentration regimes such as pseudo-first order kinetics. In addition, more complex multi-step or multi-enzyme mechanisms may be aggregated into a small number of pseudo-reactions [74, 11, 69, 110]. Often these represent empirical fits rather than a physical limit of a specific mechanism [192]. While ARLs generally simplify the set of reactions, they increase the mathematical complexity of the rate laws which, in general, are non-linear functions of substrate, product, activator, and inhibitor concentrations. Moreover, ARL based models are valid only over a particular concentration and time regime [172, 154].

An alternative modeling technique is to use mass action rate laws (MRLs). In this method, every reaction is an elementary chemical reaction [119] whose rate is described by the law of mass action [178]. These types of models have the advantage that the law of mass action is valid over all time scales and concentration regimes. However, this type of model suffers from two principal disadvantages. First, MRL can be difficult to solve numerically. Second, much of the data that exists for individual enzymes are not the elementary rate constants but rather are equilibrium binding constants [13, 40, 39].

The numerical difficulty of solving MRL models comes from two sources. First, by replacing aggregate reactions with many more elementary reactions, the number of differential equations increases. Second, MRL models tend to be stiff, which is to say that some species, such as enzyme substrate complexes, equilibrate quickly, while others, such as the production of products, proceed more slowly [68]. In oscillating systems or in enzyme cascades these fast modes can persist over the length of the simulation. Both of these problems have been greatly mitigated by improvements in stiff solvers [35, 6] and a general increase in computing power. In the previous chapter we simulated enzyme cascades with thousands of elementary reactions on conventional desktop hardware. In this work, we also used the Kronecker Bio toolkit which uses the stiff solver `ode15s` (see Figure 6-5.)

The second problem, that of limited data, is more fundamental. Much of the data available for enzymes is *in vitro* data and expressed in terms of the parameters of ARLs. For example, the BRENDA database contains $v_{\max}$, K_{M}, and K_{I} for many bacterial enzymes [40, 39]. Foundational work by King and Altman [96] and Cleland [46] provides one solution to this problem. In that work the authors describe a graphical method to derive the steady state rate law from a system of elementary reactions. This method has since been developed into a formal algorithm [49] and is available as a web tool [104]. We will refer generally to this scheme of automatically transforming a MRL into a stead state ARL as the King–Altman method. Through this method it is possible to automatically derive an ARL rate law that is the steady state limit of the system of mass action reactions. This provides a bridge between the elementary rate constants of the mass action model and the parameters of the ARL.

Some recent work has applied the King–Altman method to compute elementary rate con-

stants from an ARL model [192]. In that work, individual ARLs were converted into MRLs. The particular ARLs were Michaelis–Menten steady-state rate laws. Because of the special form of these rate laws, it was possible to relate term by term the coefficients in the original ARL implementation with the steady state expressions derived from the MRL using the King-Altman method. The parameters were then solved by constructing a hybrid model in which a single enzyme mechanism from the ARL model was replaced by its equivalent system of elementary reactions. The newly introduced parameters were fit to data generated by the intact ARL model using a dynamic optimization procedure and subject to the equality constraints introduced by the King-Altman method. The method was applied to a model of *Escherichia coli* central carbon metabolism [41]. Ultimately, this method has two major shortcomings. First, it could only be applied in situations where there was a direct term-by-term correspondence between the rate law produced by the King-Altman method and the ARL model. This is because the equality constraints used in the optimization relate particular terms in the ARL to a set of elementary rate constants. In cases where the ARL is based on an empirical rate law, or simply based on some other approximation (such as Hill binding, or pseudo-first order) the method could not be applied. Second, the fact that the parameters are estimated in the context of the remainder of the ARL greatly increased the computational cost. Together, these two facts limited the application of this method to only 9 of the 30 enzymes in the model.

Here we present a similar approach that has a broader applicability. Like the previous method, we constructed an elementary reaction implementation of for each enzyme mechanism. We then derived a steady state rate law using the King-Altman method. Finally we compared this rate law to the original ARL model. However, unlike the previous work, we compared the rate laws directly by minimizing the velocity vs. substrate curve for each enzyme. Operationally, this involved computing the steady state reaction velocity at all values of the substrate, product, inhibitor, and activator concentrations. Using this data we then computed the least squares difference between ARL model and the MRL. We then solved for the value of the elementary rate constants that minimized this difference. Once the individual enzymatic steps were parameterized, we assembled the full MRL model and compared its time dynamics to the ARL. Additional parameter fitting could have been done at this stage. However, for the

model considered here, this was not necessary.

There are several advantage of this method: First, the calculations depend only on a single enzyme at a time. As a result the the problem can be decomposed into small pieces which are easier to solve and can be solve in parallel. Second the fact that we are computing the steady-state rate curve means that the optimization procedure does not involve solving differential equations. Finally, there does not have to be a one-to-one correspondence between terms in our steady state rate law and the original ARL. This allows us to match a larger set of approximations than the previous method. Together, these advantages allowed us to convert more complicated mechanisms than was previously possible.

To demonstrate the utility of this method we have applied it to the same model of *E. coli* central carbon metabolism [41]. We demonstrate that using our method we were able to convert all 30 enzymes from the ARL original models. For each enzyme we are able to find a set of parameters such that the steady state rate law agrees well with the original ARL model. Moreover, simulations of the full MRL model agrees well with the original ARL. Note that this is despite the fact that no time course data was used to train the MRL model. This suggests that it is possible to convert ARL models into a MRL representation using our method.

6.2 Methods

6.3 Problem Formulation

We converted the aggregated rate law (ARL) model to a mass action rate law (MRL) model one enzyme at a time. This is possible because every ARL in this model is a function only of metabolite concentrations, and (implicitly) of the concentration of a single enzyme through the $v_{\max}$ parameter. These rate laws take on a variety of mathematical forms, so they are represented here as a general nonlinear function of metabolite concentrations.

$$\frac{dx_p}{dt} = -\frac{dx_s}{dt} = r_{\text{ARL}}(\mathbf{x}) \tag{6.1}$$

Where $\mathbf{x}$ is the vector of metabolite concentrations, x_s is a substrate of the enzyme, and x_p is a product. In addition to substrates and products, many rate laws are modulated by the concentration of activators and inhibitors. While in general the rate law could depend on many or all of the metabolite concentrations, in this work the most was four.

6.4 Selecting a Mass Action Mechanism

To convert the aggregated rate law (ARL) model of an enzyme to a mass action rate law (MRL) model the first step was to choose a reaction scheme comprised of elementary reactions. In most cases this mechanism was stated explicitly in the model. However, in a few cases the mechanism had to be inferred from the mathematical structure of the ARL. Next, an expression for the steady state limit of the MRL was derived using the King-Altman method. This gave a new rate law of the form:

$$r_{\mathrm{SS}} = \frac{\sum_i n_i(\mathbf{k}) \prod_j x_j^{\alpha_{i,j}}}{\sum_i d_i(\mathbf{k}) \prod_j x_j^{\beta_{i,j}}} \tag{6.2}$$

where, n_i, and d_i are a polynomials of $\mathbf{k}$ the vector of elementary rate constants $\alpha_{i,j}$ and $\beta_{i,j}$ are non-negative integers.

6.5 Solving for Elementary Rate Constants

To solve for the various elementary rate constants we optimized the elementary rate constants so as to minimize the average least squares difference between the steady state limit of the mass action rate-law and that of the aggregated rate-law:

$$\hat{\mathbf{p}} = \arg\min_{\mathbf{p}\in\mathbf{P}} \int_{\mathbf{c}_{lb}}^{\mathbf{c}_{ub}} [r_{\mathrm{ARL}}(\mathbf{c}) - r_{\mathrm{SS}}(\mathbf{c}, p)]^2 d\mathbf{c}. \tag{6.3}$$

The integral was implemented by using the trapezoidal rule with concentrations sampled at log spaced intervals. The optimization itself was implemented using MATLAB's `fmincon` with the active set algorithm [70] and complex step finite difference gradients. In a few cases, the

optimizer would get stuck in a local minima. In these cases, the pattern search algorithm [171, 107, 108, 109, 9] was used to produce an improved initial guess which was then refined using active-set.

6.5.1 Software

The elementary reactions schemes were represented as biokin files [104]. These files were processed using the biokin web-tool which produces a steady state rate law as a C–function. This function takes a vector of concentrations and a vector of elementary rate constants and returns the steady state rate. The C–function was translated into a MATLAB to allow it to be incorporated into the overall optimization procedure.

6.5.2 Under-determined Parameters

The MRLs have more parameters than the ARLs that they are fit to. As a result, the mass action parameters are underdetermined. To address this, we set some of the free parameters to physically reasonable values. Which particular parameters were set is discussed below with the individual reaction mechanism, but in general the on–rates of substrates binding to enzymes were chosen to be approximately the diffusion limit ($10^6 mM^{-1}s^{-1}$) for protein small molecule binding [60]. This is a reasonable assumption, as many small molecule substrates bind with this approximate on–rate [121]. In addition, choosing a fast on–rate helps to ensure that the MRLs approach their steady state quickly where they will match the ARLs. Ultimately, these constraints are artificial and should be fit to data from a properly designed experiment.

6.6 Results and Discussion

In this work, we converted the Chassagnole model [41] of *Escherichia coli* central carbon metabolism from aggregated rate law to mass action rate law. The original model consists of 30 Enzymes and 17 metabolites. The model does not describe the production of various energy molecules and cofactors including ATP, ADP, AMP, NAD, NADH, NADP, and NADPH. For this work, these concentrations were considered to be constants [175].

$$\emptyset \xrightarrow{k_{\text{synth}}} P$$

Scheme 6.1: Constitutive Production

$$\mathrm{S} + \mathrm{E} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \mathrm{E{:}S} \xrightarrow{k_3} \mathrm{E} + \mathrm{P}$$

Scheme 6.2: A Uni–Uni Irreversable Enzyme

Our general approach was to fit a mass action mechanism to each individual enzyme mechanism. Below we show the mass action mechanism for the various enzymes. Each mechanism was compared to the original ARL mechanism. We compared the the rate vs. concentration surface by computing the pseudo-steady state reaction rate at a range of substrate, activator, and inhibitor concentrations. Where there was more than two concentrations that affected the rate we plotted one– and two–diminutional projections of this multidimensional surface. However, the comparison used in the optimization was for the full surface. We also compared the time dynamics of the enzyme by performing a simulation starting a the model initial conditions. To test the various levels of approximation we plot the results for the ARL model, the both the full mass action model and the steady state limit of the mass action model. A representative set of these plots are shown below, and the full set of these plots are shown in Appendix B.

6.6.1 Constant Synthesis Reactions

The major source of carbon in the Chassagnole model is glucose uptake. However, pyruvate (PYR) andglyceraldehyde-3-phosphate (GAP) enter the system as the product of other cellular metabolic processes. In particular, tryptophan synthesis produces GAP and PYR and methionine synthesis produces PYR. These are modeled at constant synthesis terms with k_{synth} equal to the rate from the ARL model.

6.6.2 Uni–Uni Irreversible Reactions

A number of enzymes are modeled as a uni–uni irreversible enzyme (Table 6.1) with a Michaelis–Menten rate law:

$$r_{\mathrm{ARL}} = \frac{\mathrm{v_{max}S}/K_{\mathrm{M}}}{1 + \mathrm{S}/K_{\mathrm{M}}} \tag{6.4}$$

A mass action mechanism that can lead to this rate law is given by Scheme 6.2, where:

$$K_{\mathrm{M}} = \frac{k_2 + k_3}{k_1} \tag{6.5}$$

$$v_{\mathrm{max}} = \mathrm{E}_0 k_3. \tag{6.6}$$

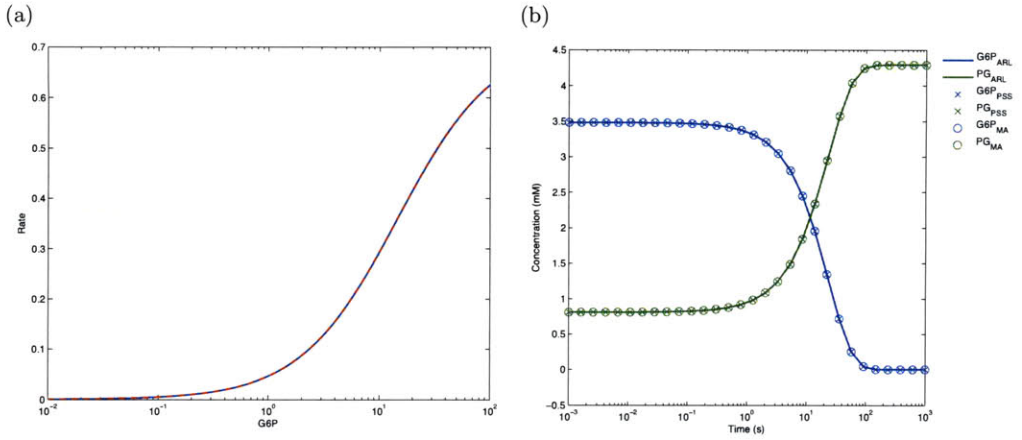
With three rate constants and one additional species, the mass-action mechanism has two degrees of freedom more than the original rate law. To prescribe these parameters we made two somewhat arbitrary choices. First, the total enzyme concentration was set to $10^{-3}mM$ which is a typical bacterial enzyme concentration [99]. Second, k_{on} was set to $10^{6}mM^{-1}s^{-1}$, which is a typical diffusion limit for protein small molecule interactions [121, 60]. Given these two choices we can solve for the remaining parameters.

$$k_1 = 10^{6}mM^{-1}s^{-1} \tag{6.7}$$

$$k_2 = K_{\mathrm{M}}k_{\mathrm{on}} - v_{\mathrm{max}}/\mathrm{E}_0 \tag{6.8}$$

$$k_3 = v_{\mathrm{max}}/\mathrm{E}_0 \tag{6.9}$$

Note that it is possible to generate a k_2 that is negative. While this did not occur for any of the enzymes in this model the solution would be to increase k_{on} such that $k_{\mathrm{on}} > v_{\mathrm{max}}/(\mathrm{E}_0 K_{\mathrm{M}})$. The result for glucose-6-phosphate-dehydrogenase is shown in Figure 6.6.2. Figure 6.6.2 (a) shows the rate vs substrate curve for the enzyme. The ARL mechanism (solid blue line) and the MRL mechanism over-plot exactly. Figure 6.6.2(b) shows a simulation for this enzyme starting from the model initial conditions for enzyme concentration, as well as the concentration of 6-phosphogluconate (6PG) and glucose-6-phosphate (G6P). The solid lines show the ARL model, the $\times$s show the steady state rate law derived by the King-Altman method, and the $\bigcirc$s show



Simulation from initial conditions.

Figure 6-1: Fit of G6PDH Mass Action Mechanism to Aggregated Rate Law mechanism. (a) Rate vs. substrate plot. (b)

Enzyme	Substrate	Product	Elaboration
ribose phosphate pyrophosphokinase	ribulose-5-phosphate	nucleotide	
glycerol 3-phosphate-dehydrogenase	dihydroxyacetone phosphate	glycerol	
serine synthesis	3-phosphoglycerate	serine	
synthesis 1	phosphoenolpyruvate	carbohydrates, mureine	
synthesis 2	pyruvate	isolucine	
glucose-6-phosphate dehydrogenase	glucose-6-phosphate	6-phosphogluconate	
6-phosphogluconate dehydrogenase	6-phosphogluconate	ribose-5-phosphate	
mureine synthesis	fructose-6-phosphate	Mureine	Pseudo-0th–order
PEP carboxylase	phosphoenolpyruvate	organic amino acids	Cooperative activation
pyruvate kinase	phosphoenolpyruvate	pyruvate	Cooperative Substrate binding Cooperative activation
pyruvate dehydrogenase	pyruvate	acetyl-coenzyme A	Cooperative substrate binding
phosphoglucoisomerase	glucose-6-phosphate	fructose-1,6-bisphosphate	Allosteric inhibition

Table 6.1: Uni–Uni Irreversable Enzymes in the *E. coli* Model.

the full MRL model. Note that even at very short time scales all three models over-plot almost exactly.

6.6.3 Variations on the Uni–Uni Irreversible Mechanism

In addition to the basic uni–uni irreversible mechanism, there are a number of related mechanisms. In general, these variations arise from the fact the enzymes contain allosteric activating and inhibitory sites. In addition, many of the enzymes exist in multi-enzyme complexes.

Pseudo–Zeroth-Order Reactions

In the Chassagnole model [41], mureine (Mur) synthesis is modeled as a zeroth–order-reaction. Essentially this means that fructose-6-phosphate (F6P) is shunted into the mureine synthesis pathway at a constant rate that is independent of any of the model species.

$$\text{F6P} \xrightarrow{k} \text{Mur} \qquad -\frac{d\text{F6P}}{dt} = \frac{d\text{Mur}}{dt} = k \tag{6.10}$$

Reactions of this type do not have the proper saturation behavior. In particular, if the concentration of F6P becomes 0 in the model it will continue to be consumed by the reaction and will become negative. In subsequent design work on this pathway, this was a persistent problem and caused many of the ARL simulations not to converge.

In reality, synthesis of Mureine (Mur) is a complex multi-step process that combines UDP-N-acetylglucosamine from fructose-6-phosphate with UDP-N-acetylmuramic acid from UDP-N-acetylglucosamine, UDP-N-acetylmuramyl-pentapeptide from UDP-N-acetylmuramic acid, and D-glutamic acid and dipeptide D-alanyl-D-alanine [12]. A more complete model of this process would have to take into account all of these steps. However, for the purposes of this work we model the first step of this synthesis pathway glucosamine-6-phosphate synthase converting F6P to glucosamine-6-phospate. We model the process as a uni–uni irreversible enzyme as described in Section 6.6.2 in the pseudo–zeroth-order limit. In this limit where $\text{S} \gg K_M$ Equation 6.4 becomes:

$$\frac{d\text{S}}{dt} \approx v_{\max} \tag{6.11}$$

which matches the given rate law 6.10.

The parameters for the MRL model are determined as with the uni–uni irreversible enzyme, with the exception that a value for K_M is not provided by the ARL. We chose to use $K_\text{M} = 0.2mM$ which is the observed K_M of glucosamine-6-phosphate synthase for F6P [51]. The results of the fit are shown in Figure 6-2.

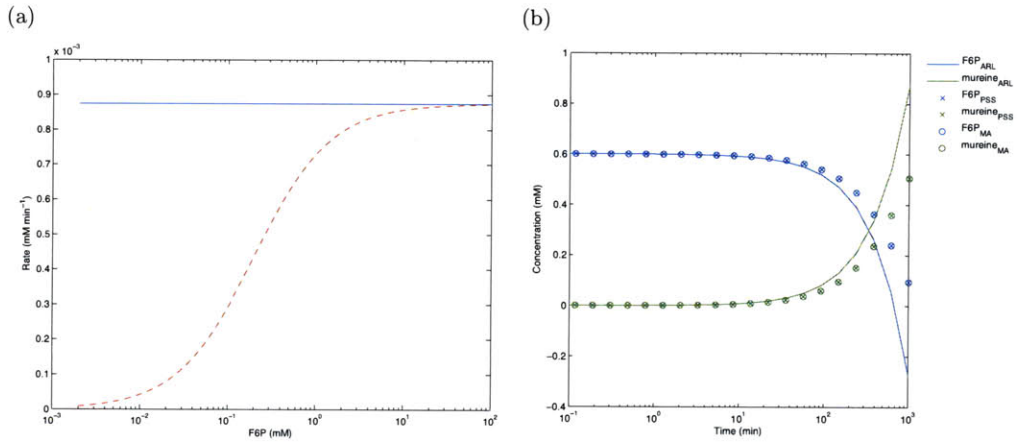


Figure 6-2: Parameter Fitting for Murine Synthesis a Pseudo–Zerorth-Order Mechanism. (a) shows the rate of production of mureine as a function of F6P concentration. The blue line shows the rate for the ARL and the red dashed line for the MRL. (b) Simulation of murine synthesis starting from the model initial conditions. Note that the ARL model (solid lines) eventually goes negative as the pool of F6P is depleted. Both the MRL model ◯ and the steady state limit of the MRL × saturates as the concentration of F6P approaches zero.

Uni–Uni Irreversible with Cooperative Activator Binding

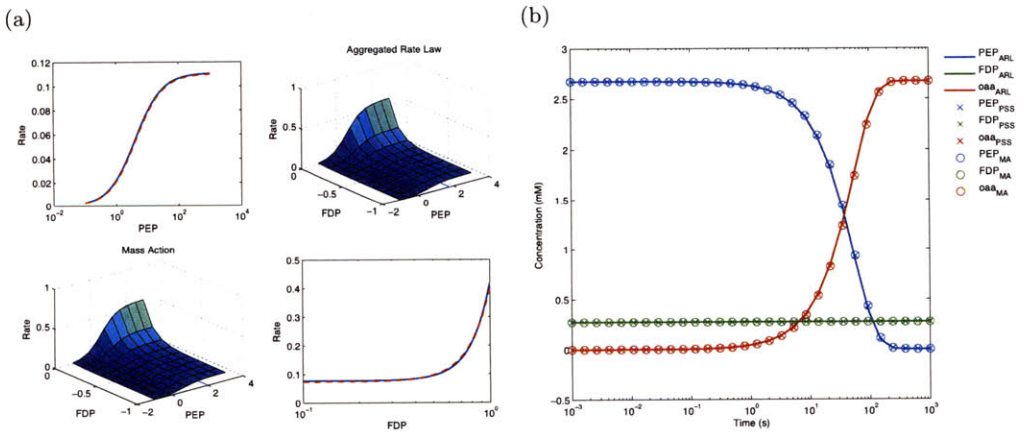
PEP carboxylase (PEPCxylase) catalyzes the reaction from phosphoenolpyruvate (PEP) to form organic amino acids. This reaction is activated by binding to an allosteric activator fructose-1,6-bisphosphate (FDP). In the Chassagnole model, the activation is modeled with fourth-order Hill binding. Scheme 6.3 shows a network of elementary reactions that can give this effect. Figure 6-3 shows the comparison between the ARL and the MRL models. The scheme shows that this mechanism can be extended to arbitrary numbers of activators. For the purposes of matching a particular Hill coefficient, the n the number of activators that can be bound must be larger than the desired Hill coefficient [154].

Uni-Uni Irreversible with Cooperative Substrate Binding

The enzyme pyruvate dehydrogenase (PDH) is one of three enzymes in the pyruvate dehydrogenase complex. Together these three enzymes convert PYR, Coenzyme A, and NAD^+ to acetyl-CoA, CO_2, and H^+. PDH catalyses the first step in this reaction, which involves the evo-

$$
\begin{array}{ccccccc}
\mathrm{E}_0 & \underset{-\mathrm{A}}{\overset{+\mathrm{A}}{\rightleftharpoons}} & \mathrm{E}_1 & \underset{-\mathrm{A}}{\overset{+\mathrm{A}}{\rightleftharpoons}} & \ldots & \underset{-\mathrm{A}}{\overset{+\mathrm{A}}{\rightleftharpoons}} & \mathrm{E}_n \\
+ & & + & & \ldots & & + \\
\mathrm{S} & & \mathrm{S} & & \ldots & & \mathrm{S} \\
\updownarrow & & \updownarrow & & \ldots & & \updownarrow \\
\mathrm{E}_0 : \mathrm{S} & & \mathrm{E}_1 : \mathrm{S} & & \ldots & & \mathrm{E}_n : \mathrm{S} \\
\downarrow & & \downarrow & & \ldots & & \downarrow \\
\mathrm{E}_0 + \mathrm{P} & & \mathrm{E}_1 + \mathrm{P} & & \ldots & & \mathrm{E}_n + \mathrm{P}
\end{array}
$$

Scheme 6.3: Mechanism of Uni–Uni Irreversible Enzyme with Cooperative Activator Binding.



Simulation from initial conditions.

Figure 6-3: Fit of PEPCxylase Mass Action Mechanism to Aggregated Rate Law mechanism. (a) Rate vs. substrate plot . (b)

lution of CO_2, making it irreversible. This is also the rate limiting step in the overall reaction performed by the pyruvate dehydrogenase complex, making PDH the key checkpoint between glycolysis and the tricarboxylic acid cycle (TCA cycle).

The pyruvate dehydrogenase complex is a multi-subunit complex that in many organisms is formed into a protein pseudo crystal. However, each subunit of this complex contains two molecules of PDH, each of which have two catalytic sites. Chassagnole models a PDH subunit as a single uni–uni irreversible enzyme with cooperative substrate binding and a Hill coefficient of 4. This simplification corresponds to the assumption that each subunit acts as a perfectly cooperative enzyme with four binding sites, and that subunits are independent of each other. Following this simplification, we model PDH using Scheme 6.4.

$$\begin{array}{ccccccc} \mathrm{E} + n\mathrm{A} & \rightleftharpoons & \mathrm{E:A_1} + (n-1)\mathrm{A} & \rightleftharpoons & \ldots & \rightleftharpoons & \mathrm{E:A}_n \\ & & \downarrow & & & & \downarrow \\ & & \mathrm{E} + (n-1)\mathrm{A} + \mathrm{P} & & & & \mathrm{E:A}_{n-1} + \mathrm{P} \end{array}$$

Scheme 6.4: Uni-Uni Irrev with Cooperative Substrate Binding

Uni-Uni Irreversible with Cooperative Binding and Cooperative Allosteric Activation

The tetrameric phosphofructokinase (PFK) is a multi-site irreversible enzyme that catalyzes the conversion F6P to FDP. Chassagnole models the substrate binding as cooperative with a Hill coefficient of 11. To achieve such a high order Hill coefficient in the MRL model the enzyme is modeled as having 11 substrate binding sites. In the ARL model, PEP binds with a Hill coefficient of 4 which is more plausible given there is one allosteric site per PFK monomer. Scheme 6.5 shows the general MRL mechanism. To agree with the Chassagnole model, we chose n of 11 and m of 4. Figure 6-4 shows that there is good agreement between the two mechanisms.

While our MRL implementation gives good agreement with the ARL model (Figure 6-4), it doesn't represent a very plausible mechanism. One possible explanation for such a high Hill coefficient for this enzyme is more likely due to product activation [85]. PEP is an allosteric

$$
\begin{array}{cccccccc}
 & + mB & \rightleftharpoons & E:B_1 + (m-1)B & \rightleftharpoons & \ldots & \rightleftharpoons & E:B_m \\
E & & & & & & & \\
 & + nA & \rightleftharpoons & E:A_1 + (n-1)A & \rightleftharpoons & \ldots & \rightleftharpoons & E:A_N \\
 & & & \downarrow & & & & \downarrow \\
 & & & E + (n-1)A + P & & & & E:A_{n-1} + P
\end{array}
$$

Scheme 6.5: Mechanism of Phosphofructokinase

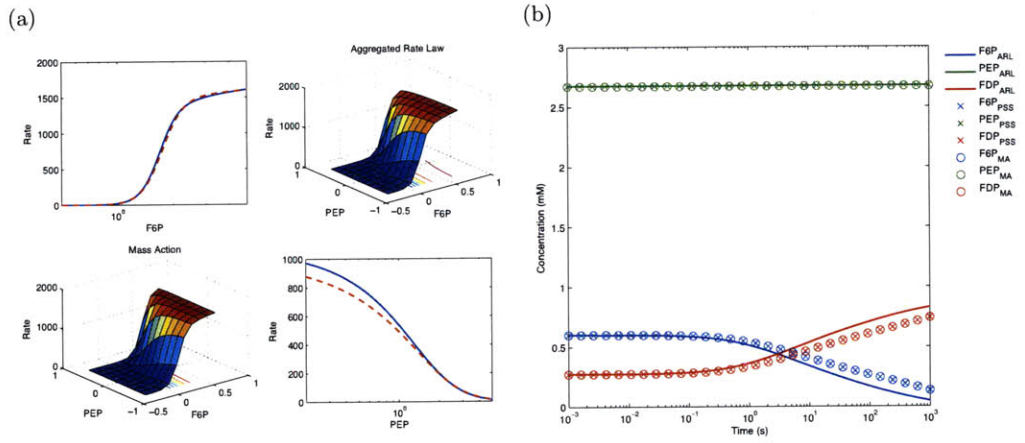


Figure 6-4: Parameter fit for Phosphofructokinase Mass Action Mechanism to Aggregated Rate Law mechanism. Rate vs substrate plot (a), and Simulation from initial conditions (b)

inhibitor of PFK.

6.6.4 Uni–Uni Reversible Reactions

A number of the reactions in the Chassagnole model are uni–uni reversible enzymes. These reactions have rate laws with the following prototype:

$$r_{\mathrm{ARL}} = \frac{\mathrm{v_{max}}(\mathrm{S} - \mathrm{P}/K_{\mathrm{eq}})}{K_{\mathrm{M}} + \mathrm{S} + \mathrm{P}/K_{\mathrm{P}}}. \tag{6.12}$$

One mass action mechanism that can give a steady state rate law of this form is:

$$\mathrm{S} + \mathrm{E} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \mathrm{E{:}S} \underset{k_3}{\overset{k_4}{\rightleftharpoons}} \mathrm{E{:}P} \underset{k_6}{\overset{k_5}{\rightleftharpoons}} \mathrm{E} + \mathrm{P}. \tag{6.13}$$

Enzyme	Substrate	Product	Elaboration
triosphospahte isomerase	dihydroxyacetone phosphate	glyceraldehyde-3-phosphate	
glyceraldehyde 3-phosphate dehydrogenase	glucose-6-phosphate	6-phosphogluconate	
phosphogylcerate kinase	1,3-diphosphosphoglycerate	3-phosphoglycerate	
phosphoglycerate mutase	3-phosphoglycerate	2-phosphoglycerate	
enolase	2-phosphoglycerate	phosphoenolpyruvate	
phosphoglucomutase	glucose-6-phosphate	glucose-1-phosphate	
phosphoglucoisomerase	glucose-6-phosphate	fructose-6-phosphate	Inhibition
ribulose phosphate isomerase	ribose-5-phosphate	xylulose-5-phosphate	Pseudo–1st-Order
ribulose phosphate epimerise	ribose-5-phosphate	ribulose-5-phosphate	Pseudo–1st-Order
glucose 1-phosphate adenyltransferase	glucose-1-phosphate	polysaccharides	Cooperative Activation
phosphofructokinase	fructose-6-phosphate	fructose-1,6-bisphosphate	Cooperative Activation
			Cooperative Inhibition

Table 6.2: Uni–Uni Reversible Enzymes in the *E. coli* Model. The table lists the enzymes along with their substrates and products, and notes any variations on the basic mechanism.

The steady state rate law (as determined by the King-Altman method) is:

$$r_{SS} = \frac{k_1 k_3 k_5 \mathrm{S} - k_2 k_4 k_6 \mathrm{P}}{(k_4 k_6 + k_2 k_6 + k_3 k_6)\mathrm{P} + (k_1 k_5 + k_1 k_4 + k_1 k_3)\mathrm{S} + (k_2 k_5 + k_2 k_4 + k_3 k_5)}. \tag{6.14}$$

As with the uni-uni irreversible mechanism, there are more rate constants in the MRL model than there are in the ARL model. As before, the enzyme concentration is assumed to be $10^-3 mM$ and the on–rates k_1 and k_6 are assumed to be $10^6 mM^{-1}s^{-1}$. Because of the simple structure, it is possible to directly solve for the elementary rate constants that match the ARL.

$$\begin{aligned}
k_1 &= 10^6 mM^{-1}s^{-1} \\
k_2 &= \frac{K_\mathrm{M} K_\mathrm{P}(\mathrm{E}_0 K_\mathrm{EQ} K_\mathrm{M} k_1 k_6 - (k_1 + K_\mathrm{EQ} k_6) v_\mathrm{max})}{\mathrm{E}_0 K_\mathrm{EQ} K_\mathrm{M} k_6 K_\mathrm{P} - (K_\mathrm{EQ} K_\mathrm{M} + K_\mathrm{P}) v_\mathrm{max}} \\
k_3 &= \frac{K_\mathrm{EQ} K_\mathrm{M} k_6 K_\mathrm{P} v_\mathrm{max}}{\mathrm{E}_0 K_\mathrm{EQ} K_\mathrm{M} k_6 K_\mathrm{P} - (K_\mathrm{EQ} K_\mathrm{M} + K_\mathrm{P}) v_\mathrm{max}} \\
k_4 &= \frac{K_\mathrm{M} k_1 K_\mathrm{P} v_\mathrm{max}}{\mathrm{E}_0 K_\mathrm{EQ} (K_\mathrm{M}^2) k_1 - (K_\mathrm{EQ} K_\mathrm{M} + K_\mathrm{P}) v_\mathrm{max}} \\
k_5 &= \frac{K_\mathrm{M} K_\mathrm{P}(\mathrm{E}_0 K_\mathrm{EQ} K_\mathrm{M} k_1 k_6 - (k_1 + K_\mathrm{EQ} k_6) v_\mathrm{max})}{\mathrm{E}_0 K_\mathrm{EQ} (K_\mathrm{M}^2) k_1 - (K_\mathrm{EQ} K_\mathrm{M} + K_\mathrm{P}) v_\mathrm{max}} \\
k_6 &= 10^6 mM^{-1}s^{-1}
\end{aligned} \tag{6.15}$$

The results for enolase (Eno) are shown in Figure 6-5. As with the uni-uni irreversible mechanism, the fits are perfect because an exact value for the elementary rate constants can be found and used as an initial guess for the optimizer.

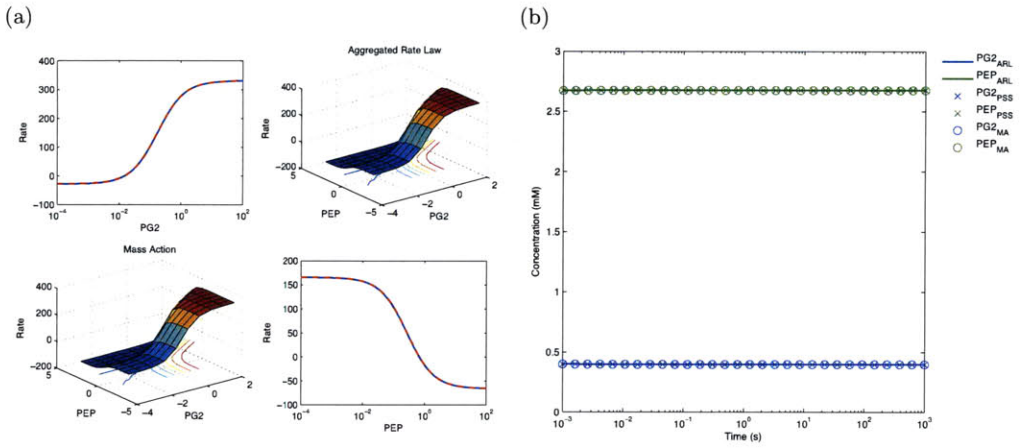


Figure 6-5: Parameter fit for Enolase a Uni–Uni Reversible Enzyme. Comparison of Mass Action Mechanism to Aggregated Rate Law Mechanism. (a)Rate vs. substrate plot. (b) Simulation from initial conditions.

Pseudo–First-Order Reversible Reactions

Chassagnole models the enzymes ribulose phosphate isomerase (Ru5P) and ribulose phosphate epimerise (R5P1) as pseudo–first-order reversible enzymes. Both rate laws have the following form:

$$r = v_{\max}(\mathrm{S} - \mathrm{P}/K_{eq}). \tag{6.16}$$

This rate law represent a limit first-order reversible mechanism at low substrate concentrations. In this limit, $K_{\mathrm{M}} \gg \mathrm{S} + \mathrm{P}/K_{\mathrm{P}}$, so the usual uni–uni reversible rate law can be approximated as:

$$r \approx \frac{v_{\max}}{K_{\mathrm{M}}}(\mathrm{S} - \mathrm{P}/K_{\mathrm{eq}}). \tag{6.17}$$

We implemented these two enzyme mechanism as uni–uni reversible enzymes. This mechanism requires two additional parameters not specified by the pseudo–first-order rate law, K_{M} and K_P. Both constants were chosen to be 10^2 mM, which is 100× greater than the maximum substrate or product concentration in the Chassagnole model.

Variations on Uni–Uni Reversible

As with the irreversible enzymes, the basic uni–uni reversible mechanism can be modified with inhibitors, activators, and cooperative binding. Table 6.2 shows the general mechanisms. These mechanisms can be quite complex, but our general approach is to model each enzyme complex explicitly. The detailed mechanisms for these rate laws can be seen in Table A.13, and the resulting fits after optimization can be seen in Appendix B. This approach greatly increases the number of parameters and the number of equations.

6.6.5 Bi–Uni Irreversible Reactions with Hill Substrate Binding

There is one bi-uni irreversible enzyme, 3-deoxy-arabino-heptulosonate-7-phosphate synthase (DHAPS) in the Chassagnole model. It catalyzes the condensation of erythrose-4-phosphate (E4P) and PEP and is the first step in the aromatic amino acid synthesis pathway. There is one isoform of DHAPS for each of the three aromatic amino acids. Each isoform performs the same reaction but has an different allosteric inhibition that is specific for one of the three aromatic amino acids. In *E. coli* the reaction is known to be sequential with PEP binding first [2]. The enzymes exist as dimers and exhibit significant cooperativity. We modeled the dimers as two bimolecular sites. The complete reaction mechanism can seen in Table A.13, and the resulting fits after optimization can be seen in Appendix B.

6.6.6 Uni–Bi Ordered Reversible Reactions

Aldolase catalyses the reversible conversion of the hexose FDP into two triose sugars dihydroxyacetone phosphate (DHAP), and GAP. The forward reaction involves the ring breaking of FDP, followed by sequential unbinding of DHAP, and then GAP. We model the reaction as a uni–bi ordered reversible enzyme. The elementary reaction scheme is shown in Scheme 6.6. The pseudo steady state rate law obtained by the King–Altman method [96, 49, 154], is shown in Equation 6.19. The resulting fits after optimization can be seen in Appendix B.

$$\mathrm{E} + \mathrm{A} \underset{k_1}{\overset{k_0}{\rightleftharpoons}} \mathrm{E{:}A} \underset{k_2}{\overset{k_3}{\rightleftharpoons}} \mathrm{E{:}XY} \underset{k_4}{\overset{k_5}{\rightleftharpoons}} \mathrm{E{:}X} + \mathrm{Q} \underset{K_6}{\overset{k_7}{\rightleftharpoons}} \mathrm{E} + \mathrm{Q} + \mathrm{P}$$

Scheme 6.6: Mechanism of Uni–Bi Ordered Reversible Enzyme

$$\begin{aligned}
n_1 &= -k_1k_3k_5k_7 \\
n_2 &= +k_0k_2k_4k_6 \\
d_1 &= k_3k_5k_7 + k_1k_5k_7 + k_2k_5k_7 \\
d_2 &= k_0k_3k_5 + k_0k_2k_5 \\
d_3 &= k_1k_3k_5 \\
d_4 &= k_1k_3k_7 + k_1k_4k_7 + k_2k_4k_7 \\
d_5 &= k_0k_3k_6 + k_0k_4k_6 + k_0k_2k_6 + k_0k_2k_4 \\
d_6 &= k_1k_3k_6 + k_1k_4k_6 + k_2k_4k_6 \\
r_{\mathrm{SS}} &= \frac{n_1\mathrm{PQ} + n_2\mathrm{A}}{d_1\mathrm{PQ} + d_2\mathrm{AQ} + d_3\mathrm{Q} + d_4\mathrm{P} + d_5\mathrm{A} + d_6}
\end{aligned} \tag{6.18}$$

6.6.7 The Full MRL Model

Once parameters were found for all of the individual enzyme mechanisms, the individual mechanism were assembled into a full pathway model. Tables A.13, A.14, and A.15 show the detailed mechanisms and parameters. To validate the model we recapitulated the experiments shown in the Chassagnole paper. First, we confirmed that the steady state of our MRL model was the same as the ARL model. Second, we reproduced the glucose pulse experiments in which a momentary injection of glucose increased the extracellular glucose concentration to 1.4 mM. Figure 6-6 shows the trajectories for the ARL model (red lines) and the MRL model (blue lines). In both cases, the models were allowed to run for 10 minutes to demonstrate the steady state. The two models performed remarkably similarly, despite the fact that all of the parameters were fit in the steady state limit of the enzyme rate laws, and did not require fine-tuning using time course data.

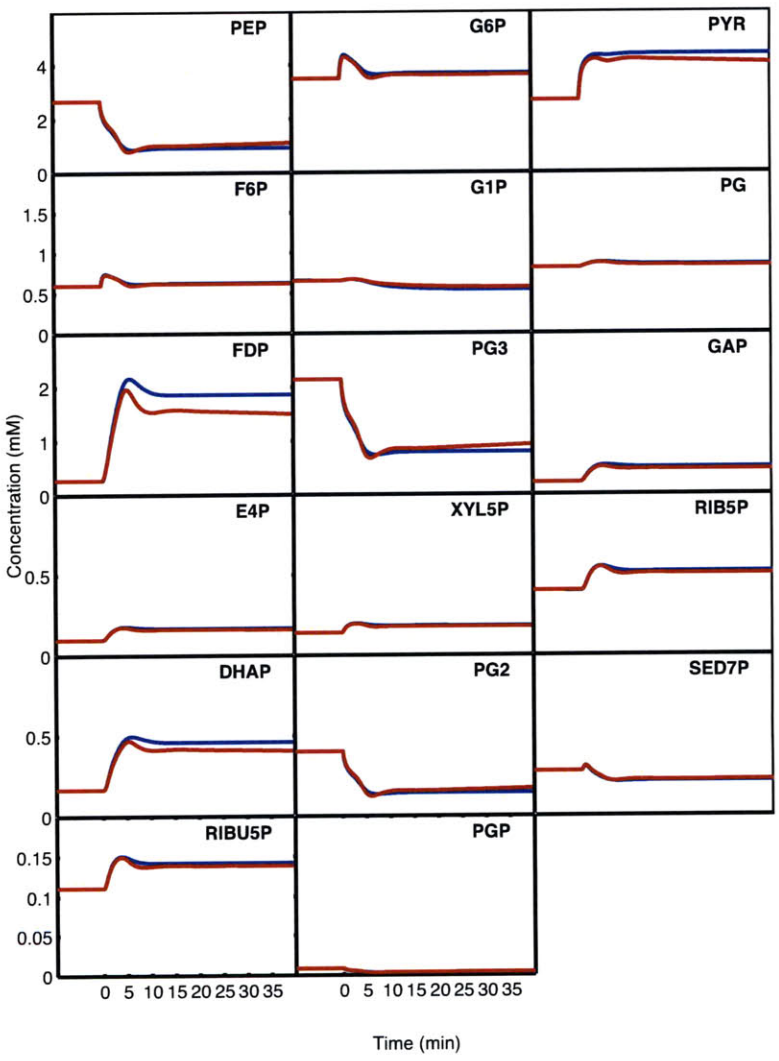


Figure 6-6: Comparison of the ARL to the MRL Model. This figure shows the response of the ARL model (red lines) and the MRL model (blue lines) to a pulse of glucose applied at $t = 0$.

In both of these simulations, we set the concentration of the co-metabolites to a constant value. To gauge the effect of this choice, we compared the two models to experimental data [41]. Figure 6-7 shows the results of this comparison. Both models match the data (diamonds) at short times (1 minute), however, the models quickly diverge from the data. In Chassagnole they correct for this by measuring the co-metabolite concentrations and then applying them as inputs to the system. This does improve the fit to the data, however, it greatly limits the applicability of the model if the time courses of co-metabolites need to be measured before the model can make a prediction. Others have used the Chassagnole model with the assumption of constant co-metabolites to good effect [175].

6.7 Discussion

We have shown here that a complex ARL model can be converted into a MRL. Moreover, we have shown that the MRL can perform similarly in the range of conditions where the ARL was developed. One limitation of both models is that the consumption and production of co-metabolites is not modeled. Simply updating the model to consume the co-metabolites (where appropriate) would not fix this problem, because it does not address the production of new high energy co-metabolites by the downstream metabolism. One way to address this would be to expand the model to include these networks, in particular, the TCA cycle and oxidative phosphorylation which would close the loop on these co-metabolites.

One possible path for implementing these reactions would be to incorporate existing models of these processes. Singh *et al.* have developed a model of the TCA cycle in *E. coli* [160] that is based on ARLs. Wu *et al.* have developed a model that includes oxidative phosphorylation in eukaryotes which could be adapted for bacteria. In addition, databases such as EcoCyc [93] and BRENDA [39] contain extensive lists of bacterial enzymes and ARL parameters.

Despite not having these reactions, it may still be possible to use this model for strain design. Vital *et al.* performed optimization on the Chassagnole model to find optimal enzyme concentrations [175]. In that work they had to assume constant co-metabolite concentrations. Within this limit it is our expectation that the MRL model will behave in a more physically

reasonable way as it leaves the concentration- and time-scales where the ARLs are valid. In particular, MRLs will saturate at both low and high substrate concentrations. It remains to be seen but we are optimistic that this will result in higher quality predictions.

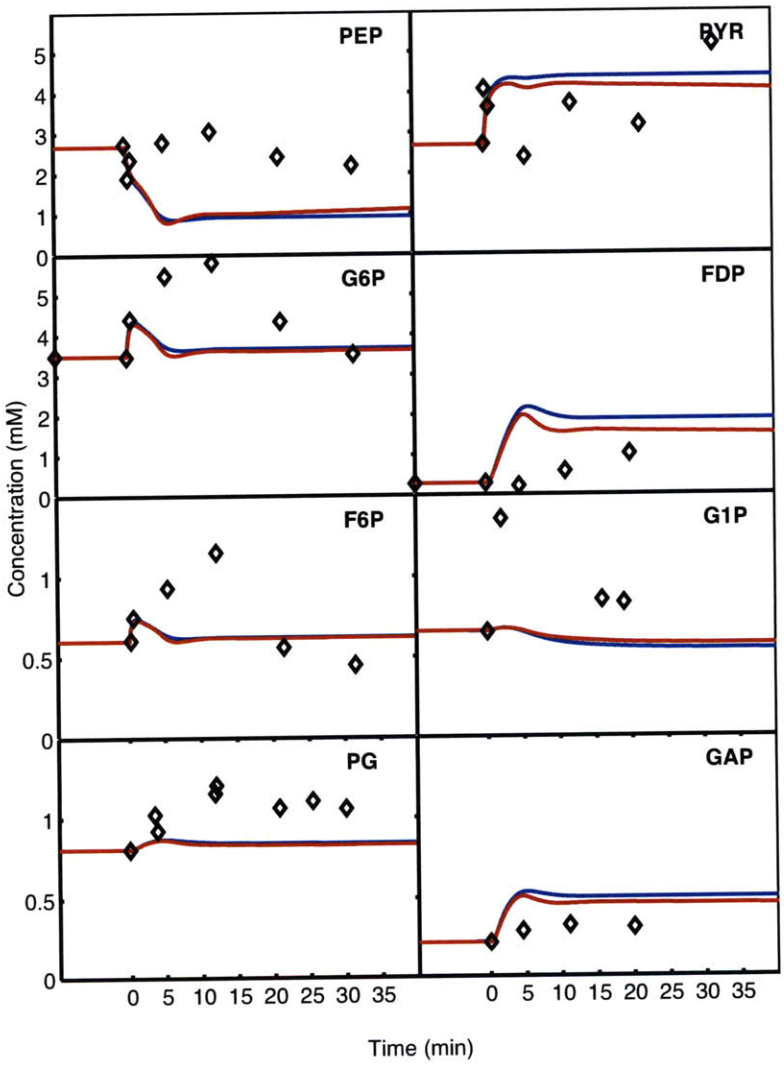


Figure 6-7: Comparison of the ARL and MRL to Experimental Time Course Data. This figure shows the fit of the ARL model (red lines) and the MRL model (blue lines) to a pulse of glucose applied at $t = 0$. In both cases the fit is rather poor after 5 minutes.

6.8 Abbreviations

ARL aggregated rate law

PSS pseudo steady state

MRL mass action rate law

TCA cycle tricarboxylic acid cycle

6.8.1 Metabolites

RIB5P ribulose-5-phosphate

G6P glucose-6-phosphate

F6P fructose-6-phosphate

G1P glucose-1-phosphate

FDP fructose-1,6-bisphosphate

DHAP dihydroxyacetone phosphate

PYR pyruvate

PEP phosphoenolpyruvate

GAP glyceraldehyde-3-phosphate

Mur Mureine

6PG 6-phosphogluconate

SED7P sedoheptulose-7-phosphate

E4P erythrose-4-phosphate

XYL5P xylulose-5-phosphate

PGP 1,3-diphosphosphoglycerate

3PG 3-phosphoglycerate

2PG 2-phosphoglycerate

RIBU5P ribose-5-phosphate

accoa acetyl-coenzyme A

6.8.2 Enzymes

PTS phosphotransferase system

PGI phosphoglucoisomerase

PFK phosphofructokinase

ALDO aldolase

TIS triosphospahte isomerase

GADPH glyceraldehyde 3-phosphate dehydrogenase

PGK phosphogylcerate kinase

PGluMu phosphoglycerate mutase

Eno enolase

PK pyruvate kinase

PDH pyruvate dehydrogenase

PEPCxylase PEP carboxylase

PGM phosphoglucomutase

G1PAT glucose 1-phosphate adenyltransferase

RPPK ribose phosphate pyrophosphokinase

G3PDH glycerol 3-phosphate-dehydrogenase

SerSynth serine synthesis

MurSynth mureine synthesis

DHAPS 3-deoxy-arabino-heptulosonate-7-phosphate synthase

TrpSynth tryptophan synthesis

MetSynth methionine synthesis

G6PDH glucose-6-phosphate dehydrogenase

PDGH 6-phosphogluconate dehydrogenase

Ru5P ribulose phosphate isomerase

R5P1 ribulose phosphate epimerise

TKa transketolase a

TKb transketolase b

TA transaldolase

Synth1 synthesis 1

Synth2 synthesis 2

Chapter 7

Conclusions and Future Directions

The promise of the genomics era and the "one gene — one disease paradigm" has largely not been realized [34]. Frustrating those efforts is the fact that disease mechanisms are complex and require a quantitative understanding to rationally develop therapies. Mechanistic dynamic models could provide a quantitative substrate to assist in this effort. One mode of analysis would be to systematically search pathways for good drug targets by computationally inhibiting every enzyme. By necessity, this will push the systems into unfamiliar and untested regions of their concentration space. Having models that are mechanistically correct and well calibrated should improve the quality of these extrapolations and make the results of such an analysis more relevant to and useful for decision-making. Despite the potential of these models, a number of challenges frustrate the construction of these high-quality models. These challenges include a lack of methods and theory to guide model formulation, model selection, and model calibration. The work presented here addresses these challenges and advanced the state of the art in experiment design for systems biology.

7.1 Dynamic Stimulation Improves Model Selection

One of the principle challenges in model construction is the fact that chemical and physical mechanisms that control biochemical processes are only partially understood. As a result, it is possible that mechanistically distinct models will fit available data and known chemistry. In

this work, we have developed a strategy for dealing with this ambiguity. This method uses the candidate modes to stimulus–response experiments, which are designed to distinguish among models with different reaction mechanisms. The results presented here show that time varying stimuli designed with these tools can allow the experimenter to distinguish among even closely related models.

We have developed two algorithms to solve these inputs. One is based on a tangent linearization, and the other is based on dynamic optimization. We evaluated our method on models of antibody–ligand binding, mitogen-activated protein kinase phosphorylation and dephosphorylation, and larger models of the epidermal growth factor receptor (EGFR) pathway. In each of these cases, a step response experiment was insufficient to distinguish between the models. We found that both design methods are able to find time varying inputs capable of distinguishing among all of the model alternatives.

One limitation of this method is that the input stimuli designed by our method could be difficult to realize experimentally. One remedy for this would be to use constraints on the input signals to restrict the optimization to the experimentally feasible space. We presented results for a monotonicity constraint which represents the situation where material can be added but not removed. Even complex inputs of this form could be generated with a computer controlled syringe pump. An alternative approach would be to use microfluidic based cell culture [77]. These systems offer very fine spatial and temporal control of the environment around cells. In particular, these systems have the potential to drive input signals up and down at high frequency. Another novel way to deliver a time varying stimulation is to use a light-switchable gene promoter system [159]. In this system, a gene can be activated or deactivated with different colors of light. One disadvantage of this approach is that the frequency of these inputs are limited by the rates of transcription and translation. However, a major advantage of this method is being able to introduce time varying inputs at multiple points in the network. Combining any of these input technologies with real-time measures of cell signaling would make for a powerful experimental platform for model building [145].

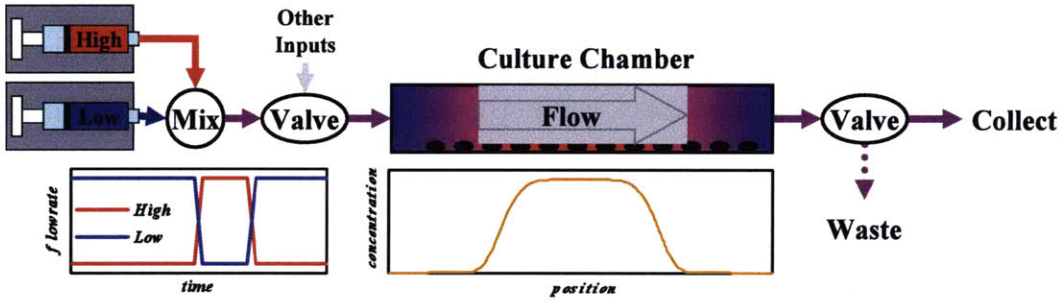


Figure 7-1: Schematic of a Microfluidics Chamber for the Delivery of Time Varying Stimulation.

7.2 Experiment Design Improves Parameter Estimation

Another modeling challenge addressed by this work is the fact that chemical kinetic models contain many unknown parameters, at least some of which must be estimated by fitting to time-course data. We examined this question in the context of a pathway model of epidermal growth factor (EGF) and neuronal growth factor (NGF) signaling. Computationally, we generated a palette of experimental perturbation data that included different doses of EGF and NGF as well as single and multiple gene knockdowns and overexpressions. While no single experiment could accurately estimate all of the parameters, we identified a set of five complementary experiments that could. These results suggest that there is reason to be optimistic about the prospects for parameter estimation in even large models.

The perturbations used in this work were all at the level of changing protein concentrations by knockdown or overexpressions. However, other experimental perturbations would also be interesting to consider, such as small molecule inhibitors. With the exception of EGF and NGF doses, all of the perturbations used in this work were static. An interesting extension to these methods would be to use inducible promoters to allow for more dynamic perturbation within the network. These types of inputs should be particularly helpful for mass action models that reactions that happen on vastly different time scales.

7.3 Experiment Design Could Improve Model Predictions

Models of biological pathways provide a substrate for understanding, manipulating, and eventually controlling complex biological processes. Many of the manipulations of these systems can be expressed in terms of changing the parameters of the system [1]. For example, overexpressing a protein can be modeled by increasing the protein's synthesis rate, or changing the protein's initial condition. The effect of a drug can be modeled by turning down the catalytic rate of drug's target enzyme. The dynamic equations of the model relate these parameter changes to the changes in the model outputs. However, as these models become more complicated, the relationships between the parameters and the model outputs becomes difficult to understand. As a result, it may not be possible to intuit the effect of changing parameter values on the model outputs by directly analyzing the model equations.

One of the main methods to address this challenge is sensitivity analysis [142, 189]. This is a numerical method in which the effect of changing a parameter is approximated by computing the derivative of the model output ψ with respect to the parameters $\mathbf{p}$ (Equation 7.1).

$$\psi(\mathbf{p}^* + \mathbf{\Delta p}) = \mathbf{p} + \mathbf{\Delta p} \left[\frac{d\psi}{d\mathbf{p}} \right|_{\mathbf{p}=\mathbf{p}^*} \tag{7.1}$$

These outputs may be the concentrations of some chemical species at a particular time, or a more complicated feature of the network [183, 182, 190]. Sensitivity analysis is local in two senses: First, it is based on a linear expansion of the full nonlinear effect of changing the parameters. This approximation will become less accurate as change in parameter ($\mathbf{\Delta p}$) becomes larger. Second, the linearization is about a nominal parameterization $\mathbf{p}^*$. If $\mathbf{p}^*$ is not accurate, then the conclusions drawn from sensitivity analysis may not be accurate even in the limit of small parameter perturbations. It is for this second reason that this method is often referred to as local sensitivity analysis.

There are many variations on the local sensitivity method that try to mitigate the importance of the nominal parameter value [142]. Global sensitivity analysis refers to a class of methods that try to characterize the parameter space in a more holistic sense. This may be

accomplished by computing the local sensitivities at many points in parameter space, by estimating higher order derivatives, or other methods [148]. While these methods tend to reduce the importance of the nominal parameterization, they also tend to emphasize sensitivities that are consistent across the entire parameter space. Where these directions exist, this can provide reliable predictions that are independent of the nominal parameter set. However, it greatly limits the predictive power of the model as only global trends are captured.

The methods we have developed in Chapter 4 have the potential to impact the application of sensitivity analysis. In this work, we used experiment design to greatly reduce the set of feasible parameters. If local sensitivity analysis is performed after our method is used to refine the parameter set, then there would be additional confidence that the nominal parameter set is close to the true parameter set. Moreover, we know that the directions where the parameter uncertainty is high correspond to directions of low sensitivity. Together, theses facts suggest that our design procedure for improving parameter estimates would also improve the accuracy of local sensitivity analysis.

7.4 EGFR Modeling: Moving Beyond Step Response Experiments

Throughout this work we have endeavored to show that mechanistic modeling combined with experiment design is a powerful combination for understanding systems biology. The work has demonstrated a real need to improve models of growth factor signaling. However, the results presented here are mostly methods development and proof of concept studies. As a future direction, it is our hope that these methods will be applied to this important signaling pathway.

Much of our understanding of cell signaling comes from steady state and step response experiments. This imposes a kind of bias on the way we think about signaling biology and the importance of particular mechanisms. In Chapter 3, we showed that even widely used models may not accurately predict the response to perturbations other than a single step change in the input. Many physiological signals vary with time, and drug regimens are often delivered on a

periodic schedule. As a result, this step response bias may ultimately limit the ability of these models to make predictions that go beyond the laboratory.

In this work we tested the idea that pulses of EGF could help elucidate the relative contribution of different feedback loops within the EGFR network. The idea behind this stimulus is that the first pulse activates the network, including all of the feedback loops. We expect that as the response to the first pulse evolves, different feedback mechanisms will increase or decrease in importance. The second pulse then probes the state of the network at a later time. By comparing the response of the network to a second pulse delivered at various times, we hoped to be able to detect the timescale and relative importance of the various feedback mechanisms. We believe that this approach could be very general and have applications in understanding multiply connected feedback mechanisms that are common in cell signaling. In addition, these methods could be useful in understanding crosstalk between pathways. For example, the first pulse could be with one ligand and the second pulse with another. Finding waiting times where the first pulse interferes with the response to the second pulse could be an effective way of finding points of crosstalk.

These experimental results suggest that models from the literature do not accurately represent the relative strength of the various feedback loops in this pathway. In particular, we observed that the endocytosis and feedback loop was less strong than predicted by models from the literature [86, 43], and that other feedback mechanisms were likely necessary to deactivate ERK after EGF stimulation. A possible next step would be to repeat the pulse experiments presented in Chapter 3 at shorter and longer waiting times. This should give a better sense for the time scale over which the pathway resets and various feedback loops become active. In addition, various drugs that block broad biological functions could be used to help elucidate the physical mechanisms that underlie this process. One example would be to treat the cells with the endocytosis inhibitor phenylarsine oxide. Another possibility would be to use a tethered ligand that could not be endocytose.

The experimental results also suggest that the response of the experimental system is highly sensitive to the presence of other signaling molecules. Even a small dose of insulin appeared to radically change the network response to EGF. Often times, biologists talk about various cues

causing the network to rewire. This implies that the basic biochemical processes have been altered or changed. These results motivate a different interpretation. Instead of the network rewiring, environmental cues can modulate the weights of the various feedback mechanisms. This type of effect could explain the observed patient-to-patient variability observed in drugs that target molecules in this pathway.

7.5 A Standardized Formulation of Chemical Kinetics Leads to Reusable Methods and Software

Currently, there is no standard formulation for chemical kinetic models of biological signaling. Here we have developed a general and concise formulation of mass action kinetics based on sparse matrices and Kronecker products. This formulation allowed us to represent any mass action model and its partial derivatives with simple matrix equations. Based on this formalism we built several numerical methods, including the design methods presented in this work. We show the generality of this formulation by converting models that use other rate laws such as Michaelis–Menten can be converted to our formulation. We demonstrate this by converting a model of *Escherichia coli* central carbon metabolism to be fully mass action. The dynamics of the new model is similar to the original model. However, we argue that because our model is based on fewer approximations, it has the potential to be more accurate over a wider range of conditions. In addition, using the same methods we can automatically convert mass action mechanisms into the corresponding aggregated rate law mechanisms. Because the reactions are elementary reactions, the same systems of equations could be simulated in a stochastic framework. This places the mass action models in a unique position to bridge the gap between more abstracted, and more detailed simulation methods.

More important than any performance considerations was the fact that standardizing on a formulation of chemical kinetics allowed us to develop methods and code without regard for the details of any particular model. This greatly improved our productivity, and allowed for the development of a diverse set of tools. It is our hope that this development continues.

7.6 Biological Robustness Does Not Imply Parameters or Mechanism are Unknowable

A strange consequence of the fact that standard step–response experiments are poor at distinguishing among multiple mechanism and parameter sets is that there is a sense in the field that detailed mechanisms and accurate parameters are unimportant and unknowable from reasonable amounts of data. This makes a kind of sense. If it were really true that no experiment could distinguish two mechanism, then it seems unlikely that the choice of mechanism would have any impact on a prediction.

The notion that multiple mechanism and parameter sets are somehow equivalent is often conflated with the idea of biological robustness. A sketch of the argument, as it applies to parameters, is as follows: from Barkai and Liebler's foundational work [11] we know that biological systems are robust to parametric variation. From Sethn'a work [78] we know that it is impossible to know most if not all of the parameters of a network. Therefore, the exact values of parameters do not matter. The work presented here shows that the second statement is wrong. It is possible to determine the parameters of a network to high accuracy. The problem with the first statement is more subtle. Barkai and Liebler showed that certain overall features of a network could be robust to parametric variation. In the case of bacterial chemotaxis, the ability of the network to adapt perfectly was robust to parameter variation. This feature is a requirement for the biological function of the network, sensing gradients. So the parametric robustness captures a kind of functional robustness or robustness of phenotype. However, if you were to plot two trajectories of the signaling molecules of the network at two different parameterizations, they might look quite different. Moreover, even global network properties such as the adaptation time were shown to be sensitive to parameter changes. The fact that some features of the network are robust to parameter changes does not imply that different parameterizations are indistinguishable or unimportant.

So how is it that models that exist today, which are poorly calibrated and contain incomplete or wrong mechanisms, can ever make good predictions? Sethna's point made more precisely provides one possible answer. His work showed that for any particular prediction only a subset

of the parameters were important. If that subset was well calibrated by the initial experiments, then the prediction is likely to be good. The relationship between model calibration and prediction accuracy is discussed in detail in [16]. This result suggests a potential variation on our method. If somehow you knew ahead of time that the predictions you wanted to make were only sensitive to certain parameter directions, then when you were choosing experiments to define the parameters you could select the experiments that covered the subspace of parameter directions that were important for your problem. A concrete example of this is that if you knew certain proteins were never expressed at a level above the K_{M} of any of the enzymes that modify them, you could decide to ignore the parameter directions that pointed in the $(+\log k_{\mathrm{cat}}, +\log K_{\mathrm{M}})$ direction. However, the results presented here suggest that it may not be worth optimizing on the subspace of parameters to determine. Especially since a small number of experiments was able to cover the entire parameter space. In addition, if the models are to be used in systematic computational experiments, such as computational screening for drug targets or designing bacterial strains for bioreactors, it is likely that most if not all of the parameters will be important for some prediction of the design.

7.7 Experiment Design Can Improve the Quality of Models

The work presented here has shown that experiment design can be used to improve the quality of models. We have shown this in terms of selecting the best possible mechanisms, and we have shown this in terms of improving parameter accuracy. Taken together, the work presented here demonstrates that experimental design can be successfully used to improve the quality of mechanism-based chemical kinetic models.

The focus of this work has been on experiment design. However, the methods developed in this thesis have potential applications to therapeutic design. For example, instead of using the control methods for experiment design, they could be used as the basis for controlling a biological system. One way to achieve this is to pick a target function that corresponds to a desired biological outcome. The choice of target function is essentially arbitrary. In Figure 7-2, we show that the same controller can cause the output (doubly phosphorylated ERK) to spell

the word *MIT*. While this is a contrived example, one could imagine a situation where the input is the concentration of a drug and the output is a particular physiological process. In fact, the emerging field of chronotherapeutics [141] is one attempt to do just that. Much of modern drug delivery has focused on the delivery of constant (with respect to time) dose of bioactive agents. However, asthma, arthritis, duodenal ulcer, cancer, diabetes, cardiovascular disease, hypercholesterolemia, and neuronal disorders have all been shown to have a temporal component to their pathophysiology and to respond differentially to different time courses of treatment [187]. A new generation of rapid release drug delivery technologies are creating the possibility of delivering more complex time varying drug stimuli [94].

If these methods were to be used in therapies, issues of robustness would become critical. One modification to the method detailed here would be to incorporate robust optimization. This could account for uncertainty in the underlying model, the parameters, or even the parameters of the clinical intervention. Even within the confines of experiment design, the use of robustness metrics could improve the practical utility of the designed experiments.

Increasingly, synthetic biologists have turned to mechanistic models to help in design, optimization, and analysis of novel biological systems [54]. As with any model based engineering, the quality of the models directly impacts their ability to guide decision-making. The techniques presented here could have a direct impact on high quality engineering models. In addition the optimization methods presented here could be easily adapted to synthetic biology design questions such as the optimal parameterization for a biological part. As was the case with the therapy design, the incorporation of robustness metrics would make these methods even more useful to the synthetic biology community.

Ultimately, the success or failure of any of these applications is greatly influenced by the data that is used to construct, calibrate, and validate the models that support them. The methods, practices, and experiments developed here provide a set of tools to improve the model building processes, and promote the highest possible model quality.

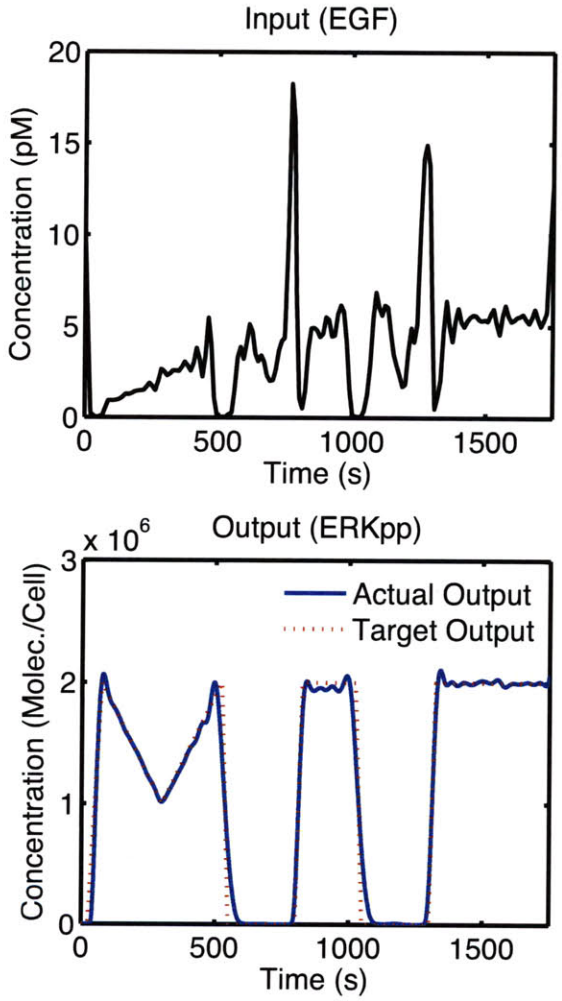


Figure 7-2: Using the Dynamic Optimization Controller to Spell out *MIT* in Phosphorylated ERK.

Appendix A

Kronecker Models

A.1 Models of Antibody–Ligand Binding

Table A.1: One Step Antibody–Ligand Model Reactions

Reactant 1	Reactant 2	Product	k_{forward}	k_{reverse}
A	L	C	k1+	k1-

Table A.2: One Step Antibody–Ligand Model Parameters.

Parameter	Value
k1+	1.30×10^{-03}
k1-	9.92×10^{-05}

Table A.3: One Step Antibody–Ligand Model Species.

Species	Initial Concentration	Is Input
L	0	yes
P	1	no
C	0	no

Table A.4: Two Step Antibody–Ligand Model Reactions

Reactant 1	Reactant 2	Product	k_{forward}	k_{reverse}
L	P	C*	k1+	k1-
C*	$\emptyset$	C	k2+	k2-

Table A.5: Two Step Antibody Ligand Model Parameters.

Parameter	Value
k1+	1.20×10^{-02}
k1-	1.20×10^{-02}
k2+	1.30×10^{-03}
K2-	1.10×10^{-04}

Table A.6: Two Step Antibody–Ligand Model Species.

Species	Initial Concentration	Is Input
L	0	yes
P	1	no
C*	0	no
C	0	no

A.2 A Model EGFR Signaling With Additional Recycling Reactions

Table A.7: Expanded EGFR model Reactions.

Reactant 1	Reactant 2	Product	k_{forward}	k_{reverse}
GAP	∅	∅	kd214	k200
Grb2	∅	∅	kd222	k200
Sos	∅	∅	kd224	k200
Ras-GDP	∅	∅	kd226	k200
Shc	∅	∅	kd231	k200
Grb2-Sos	∅	∅	kd230	k200
EGFRi	∅	∅	k60	kd60
EGFi	∅	∅	k61	kd61
(EGF-EGFRi*)2	∅	∅	k60	kd60
(EGF-EGFRi*)2-GAP	∅	∅	k60	kd60
(EGF-EGFRi*)2-GAP-Grb2	∅	∅	k60	kd60
(EGF-EGFRi*)2-GAP-Grb2-Sos	∅	∅	k60	kd60
(EGF-EGFRi*)2-GAP-Grb2-Sos-Ras-GDP	∅	∅	k60	kd60
(EGF-EGFRi*)2-GAP-Grb2-Sos-Ras-GTP	∅	∅	k60	kd60
(EGF-EGFRi*)2-GAP-SHC	∅	∅	k60	kd60
(EGF-EGFRi*)2-GAP-SHC*	∅	∅	k60	kd60
(EGF-EGFRi*)2-GAP-SHC*-Grb2	∅	∅	k60	kd60
(EGF-EGFRi*)2-GAP-SHC*-Grb2-Sos	∅	∅	k60	kd60
(EGF-EGFRi*)2-GAP-SHC*-Grb2-Sos-Ras-GDP	∅	∅	k60	kd60
(EGF-EGFRi*)2-GAP-SHC*-Grb2-Sos-Ras-GTP	∅	∅	k60	kd60
phosphatase3	ERK-PP	ERK-PP-phosphatase3	k56	kd56
phosphatase3	ERK-P	ERK-PP-phosphatase3	k57	kd57
phosphatase3	ERK-P	ERK-P-phosphatase3	k58	kd58
phosphatase3	ERK	ERK-P-phosphatase3	k57	kd57
phosphatase3	ERKi-PP	ERKi-PP-phosphatase3	k56	kd56

Continued on next page

Table A.7 – continued from previous page

Reactant 1	Reactant 2	Product	k_{forward}	k_{reverse}
phosphatase3	ERKi-P	ERKi-PP-phosphatase3	k57	kd57
phosphatase3	ERKi-P	ERKi-P-phosphatase3	k58	kd58
phosphatase3	ERK	ERKi-P-phosphatase3	k57	kd57
MEK-PP	ERK	ERK-MEK-PP	k52	kd44
MEK-PP	ERK-P	ERK-MEK-PP	k53	kd53
MEK-PP	ERK-P	ERK-P-MEK-PP	k52	kd44
ERK-PP	MEK-PP	ERK-P-MEK-PP	k55	kd55
MEKi-PP	ERK	ERKi-MEKi-PP	k52	kd44
MEKi-PP	ERKi-P	ERKi-MEKi-PP	k53	kd53
MEKi-PP	ERKi-P	ERKi-P-MEKi-PP	k52	kd44
ERKi-PP	MEKi-PP	ERKi-P-MEKi-PP	k55	kd55
(EGF-EGFR*)2-GAP-Grb2	Prot	(EGF-EGFR*)2-GAP-Grb2-Prot	k4	kd4
(EGF-EGFRi*)2-GAP-Grb2	Proti	(EGF-EGFR*)2-GAP-Grb2-Prot	k5	kd5
EGFR	∅	EGFRi	k6	kd6
(EGF-EGFR*)2	∅	(EGF-EGFRi*)2	k6	kd6
(EGF-EGFR*)2-GAP-Grb2	∅	(EGF-EGFRi*)2-GAP-Grb2	k6	kd6
Proti	∅	Prot	k15	kd15
(EGF-EGFR*)2-GAP	∅	(EGF-EGFRi*)2-GAP	k6	kd6
(EGF-EGFR*)2-GAP-SHC	∅	(EGF-EGFRi*)2-GAP-SHC	k6	kd6
(EGF-EGFR*)2-GAP-SHC*	∅	(EGF-EGFRi*)2-GAP-SHC*	k6	kd6
(EGF-EGFR*)2-GAP-Grb2-Sos	∅	(EGF-EGFRi*)2-GAP-Grb2-Sos	k6	kd6
(EGF-EGFR*)2-GAP-Grb2-Sos	Prot	(EGF-EGFR*)2-GAP-Grb2-Sos-Prot	k4	kd4
(EGF-EGFRi*)2-GAP-Grb2-Sos	Proti	(EGF-EGFR*)2-GAP-Grb2-Sos-Prot	k5	kd5
(EGF-EGFR*)2-GAP-Grb2-Sos-Ras-GDP	∅	(EGF-EGFRi*)2-GAP-Grb2-Sos-Ras-GDP	k6	kd6
(EGF-EGFR*)2-GAP-Grb2-Sos-Ras-GDP	Prot	(EGF-EGFR*)2-GAP-Grb2-Sos-Ras-GDP-Prot	k4	kd4
(EGF-EGFRi*)2-GAP-Grb2-Sos-Ras-GDP	Proti	(EGF-EGFR*)2-GAP-Grb2-Sos-Ras-GDP-Prot	k5	kd5
(EGF-EGFR*)2-GAP-Grb2-Sos-Ras-GTP	∅	(EGF-EGFRi*)2-GAP-Grb2-Sos-Ras-GTP	k6	kd6
(EGF-EGFR*)2-GAP-Grb2-Sos-Ras-GTP	Prot	(EGF-EGFR*)2-GAP-Grb2-Sos-Ras-GTP-Prot	k4	kd4
(EGF-EGFRi*)2-GAP-Grb2-Sos-Ras-GTP	Proti	(EGF-EGFR*)2-GAP-Grb2-Sos-Ras-GTP-Prot	k5	kd5
(EGF-EGFR*)2-GAP-SHC*-Grb2	∅	(EGF-EGFRi*)2-GAP-SHC*-Grb2	k6	kd6
(EGF-EGFR*)2-GAP-SHC*-Grb2	Prot	(EGF-EGFR*)2-GAP-SHC*-Grb2-Prot	k4	kd4
(EGF-EGFRi*)2-GAP-SHC*-Grb2	Proti	(EGF-EGFR*)2-GAP-SHC*-Grb2-Prot	k5	kd5
(EGF-EGFR*)2-GAP-SHC*-Grb2-Sos	∅	(EGF-EGFRi*)2-GAP-SHC*-Grb2-Sos	k6	kd6
(EGF-EGFR*)2-GAP-SHC*-Grb2-Sos	Prot	(EGF-EGFR*)2-GAP-SHC*-Grb2-Sos-Prot	k4	kd4
(EGF-EGFRi*)2-GAP-SHC*-Grb2-Sos	Proti	(EGF-EGFR*)2-GAP-SHC*-Grb2-Sos-Prot	k5	kd5
(EGF-EGFR*)2-GAP-SHC*-Grb2-Sos-Ras-GDP	∅	(EGF-EGFRi*)2-GAP-SHC*-Grb2-Sos-Ras-GDP	k6	kd6
(EGF-EGFR*)2-GAP-SHC*-Grb2-Sos-Ras-GDP	Prot	(EGF-EGFR*)2-GAP-SHC*-Grb2-Sos-Ras-GDP-Prot	k4	kd4
(EGF-EGFRi*)2-GAP-SHC*-Grb2-Sos-Ras-GDP	Proti	(EGF-EGFR*)2-GAP-SHC*-Grb2-Sos-Ras-GDP-Prot	k5	kd5
(EGF-EGFR*)2-GAP-SHC*-Grb2-Sos-Ras-GTP	∅	(EGF-EGFRi*)2-GAP-SHC*-Grb2-Sos-Ras-GTP	k6	kd6
(EGF-EGFR*)2-GAP-SHC*-Grb2-Sos-Ras-GTP	Prot	(EGF-EGFR*)2-GAP-SHC*-Grb2-Sos-Ras-GTP-Prot	k4	kd4
(EGF-EGFRi*)2-GAP-SHC*-Grb2-Sos-Ras-GTP	Proti	(EGF-EGFR*)2-GAP-SHC*-Grb2-Sos-Ras-GTP-Prot	k5	kd5
phosphatase2	MEK-PP	MEK-PP-phosphatase2	k48	kd48
phosphatase2	MEK-P	MEK-PP-phosphatase2	k49	kd49
phosphatase2	MEK-P	MEK-P-phosphatase2	k50	kd50
phosphatase2	MEK	MEK-P-phosphatase2	k49	kd49
phosphatase2	MEKi-PP	MEKi-PP-phosphatase2	k48	kd48
phosphatase2	MEKi-P	MEKi-PP-phosphatase2	k49	kd49
phosphatase2	MEKi-P	MEKi-P-phosphatase2	k50	kd50
phosphatase2	MEK	MEKi-P-phosphatase2	k49	kd49
(EGF-EGFR*)2-GAP-Grb2-Sos	ERK-PP	(EGF-EGFR*)2-GAP-Grb2-Sos-ERK-PP	k126	kd126
(EGF-EGFRi*)2-GAP-Grb2-Sos	ERKi-PP	(EGF-EGFRi*)2-GAP-Grb2-Sos-ERKi-PP	k126	kd126
(EGF-EGFR*)2-GAP-SHC*-Grb2-Sos	ERK-PP	(EGF-EGFR*)2-GAP-SHC*-Grb2-Sos-ERK-PP	k126	kd126

Continued on next page

Table A.7 – continued from previous page

Reactant 1	Reactant 2	Product	k_{forward}	k_{reverse}
(EGF-EGFRi*)2-GAP-SHC*-Grb2-Sos	ERKi-PP	(EGF-EGFRi*)2-GAP-SHC*-Grb2-Sos-ERKi-PP	k126	kd126
ERK-PP	Sos	Sos-ERK-PP	k126	kd126
ERKi-PP	Sos	Sos-ERKi-PP	k126	kd126
ERK-PP	∅	(EGF-EGFR*)2-GAP-Grb2-Sos-ERK-PP	k127	kd127
ERK-PP	∅	(EGF-EGFR*)2-GAP-SHC*-Grb2-Sos-ERK-PP	k127	kd127
ERK-PP	Sosi	Sos-ERK-PP	k127	kd127
ERKi-PP	∅	(EGF-EGFRi*)2-GAP-Grb2-Sos-ERKi-PP	k127	kd127
ERKi-PP	∅	(EGF-EGFRi*)2-GAP-SHC*-Grb2-Sos-ERKi-PP	k127	kd127
ERKi-PP	Sosi	Sos-ERKi-PP	k127	kd127
Phosphatase1	Raf*	Raf*-phosphatase1	k42	kd42
Phosphatase1	Raf	Raf*-phosphatase1	k43	kd43
Phosphatase1	Rafi*	Rafi*-phosphatase1	k42	kd42
Phosphatase1	Raf	Rafi*-phosphatase1	k43	kd43
Raf*	MEK	MEK-Raf*	k44	kd52
MEK-P	Raf*	MEK-Raf*	k45	kd45
MEK-P	Raf*	MEK-P-Raf*	k44	kd52
MEK-PP	Raf*	MEK-P-Raf*	k47	kd47
Rafi*	MEK	MEK-Rafi*	k44	kd52
MEKi-P	Rafi*	MEK-Rafi*	k45	kd45
MEKi-P	Rafi*	MEK-P-Rafi*	k44	kd52
MEKi-PP	Rafi*	MEK-P-Rafi*	k47	kd47
Ras-GTP	Raf	Raf-Ras-GTP	k28	kd28
Ras-GTP*	Raf*	Raf-Ras-GTP	k29	kd29
Ras-GTPi	Raf	Raf-Ras-GTPi	k28	kd28
Ras-GTPi*	Rafi*	Raf-Ras-GTPi	k29	kd29
(EGF-EGFR*)2-GAP-Grb2-Sos	Ras-GDP	(EGF-EGFR*)2-GAP-Grb2-Sos-Ras-GDP	k18	kd18
(EGF-EGFR*)2-GAP-Grb2-Sos	Ras-GTP	(EGF-EGFR*)2-GAP-Grb2-Sos-Ras-GDP	k19	kd19
(EGF-EGFR*)2-GAP-SHC*-Grb2-Sos	Ras-GDP	(EGF-EGFR*)2-GAP-SHC*-Grb2-Sos-Ras-GDP	k18	kd18
(EGF-EGFR*)2-GAP-SHC*-Grb2-Sos	Ras-GTP	(EGF-EGFR*)2-GAP-SHC*-Grb2-Sos-Ras-GDP	k19	kd19
(EGF-EGFRi*)2-GAP-Grb2-Sos	Ras-GDP	(EGF-EGFRi*)2-GAP-Grb2-Sos-Ras-GDP	k18	kd18
(EGF-EGFRi*)2-GAP-Grb2-Sos	Ras-GTPi	(EGF-EGFRi*)2-GAP-Grb2-Sos-Ras-GDP	k19	kd19
(EGF-EGFRi*)2-GAP-SHC*-Grb2-Sos	Ras-GDP	(EGF-EGFRi*)2-GAP-SHC*-Grb2-Sos-Ras-GDP	k18	kd18
(EGF-EGFRi*)2-GAP-SHC*-Grb2-Sos	Ras-GTPi	(EGF-EGFRi*)2-GAP-SHC*-Grb2-Sos-Ras-GDP	k19	kd19
(EGF-EGFRi*)2	GAP	(EGF-EGFRi*)2-GAP	k8	kd8
(EGF-EGFR*)2-GAP-Grb2-Sos	Ras-GTP*	(EGF-EGFR*)2-GAP-Grb2-Sos-Ras-GTP	k20	kd20
(EGF-EGFR*)2-GAP-Grb2-Sos	Ras-GDP	(EGF-EGFR*)2-GAP-Grb2-Sos-Ras-GTP	k21	kd21
(EGF-EGFR*)2-GAP-SHC*-Grb2-Sos	Ras-GTP*	(EGF-EGFR*)2-GAP-SHC*-Grb2-Sos-Ras-GTP	k20	kd20
(EGF-EGFR*)2-GAP-SHC*-Grb2-Sos	Ras-GDP	(EGF-EGFR*)2-GAP-SHC*-Grb2-Sos-Ras-GTP	k21	kd21
(EGF-EGFRi*)2-GAP-Grb2-Sos	Ras-GTPi*	(EGF-EGFRi*)2-GAP-Grb2-Sos-Ras-GTP	k20	kd20
(EGF-EGFRi*)2-GAP-Grb2-Sos	Ras-GDP	(EGF-EGFRi*)2-GAP-Grb2-Sos-Ras-GTP	k21	kd21
(EGF-EGFRi*)2-GAP-SHC*-Grb2-Sos	Ras-GTPi*	(EGF-EGFRi*)2-GAP-SHC*-Grb2-Sos-Ras-GTP	k20	kd20
(EGF-EGFRi*)2-GAP-SHC*-Grb2-Sos	Ras-GDP	(EGF-EGFRi*)2-GAP-SHC*-Grb2-Sos-Ras-GTP	k21	kd21
EGFR	EGF	EGF-EGFR	k1	kd1
EGF-EGFR	EGF-EGFR	(EGF-EGFR)2	k2	kd2
(EGF-EGFR)2	∅	(EGF-EGFR*)2	k3	kd3
EGFRi	EGFi	EGF-EGFRi	k10b	kd10
EGF-EGFRi	EGF-EGFRi	(EGF-EGFRi)2	k2	kd2
(EGF-EGFRi)2	∅	(EGF-EGFRi*)2	k3	kd3
∅	∅	EGFR	k13	kd13
(EGF-EGFR*)2	GAP	(EGF-EGFR*)2-GAP	k8	kd8
(EGF-EGFR*)2-GAP	Grb2	(EGF-EGFR*)2-GAP-Grb2	k16	kd63
(EGF-EGFR*)2-GAP-Grb2	Sos	(EGF-EGFR*)2-GAP-Grb2-Sos	k17	kd17

Continued on next page

Table A.7 – continued from previous page

Reactant 1	Reactant 2	Product	k_{forward}	k_{reverse}
(EGF-EGFR*)2-GAP	Shc	(EGF-EGFR*)2-GAP-SHC	k22	kd22
(EGF-EGFR*)2-GAP-SHC	∅	(EGF-EGFR*)2-GAP-SHC*	k23	kd23
(EGF-EGFR*)2-GAP-SHC*	Grb2	(EGF-EGFR*)2-GAP-SHC*-Grb2	k16	kd24
(EGF-EGFR*)2-GAP-SHC*-Grb2	Sos	(EGF-EGFR*)2-GAP-SHC*-Grb2-Sos	k25	kd25
(EGF-EGFR*)2-GAP	Shc*-Grb2-Sos	(EGF-EGFR*)2-GAP-SHC*-Grb2-Sos	k32	kd32
Grb2-Sos	Shc*	Shc*-Grb2-Sos	k33	kd33
(EGF-EGFR*)2-GAP	Grb2-Sos	(EGF-EGFR*)2-GAP-Grb2-Sos	k34	kd34
Grb2	Sos	Grb2-Sos	k35	kd35
Shc*	∅	Shc	k36	kd36
(EGF-EGFR*)2-GAP	Shc*	(EGF-EGFR*)2-GAP-SHC*	k37	kd37
Grb2	Shc*	Shc*-Grb2	k16	kd24
(EGF-EGFR*)2-GAP	Shc*-Grb2	(EGF-EGFR*)2-GAP-SHC*-Grb2	k37	kd37
Shc*-Grb2	Sos	Shc*-Grb2-Sos	k40	kd40
(EGF-EGFR*)2-GAP-SHC*	Grb2-Sos	(EGF-EGFR*)2-GAP-SHC*-Grb2-Sos	k41	kd41
(EGF-EGFRi*)2-GAP	Grb2	(EGF-EGFRi*)2-GAP-Grb2	k16	kd63
(EGF-EGFRi*)2-GAP-Grb2	Sos	(EGF-EGFRi*)2-GAP-Grb2-Sos	k17	kd17
(EGF-EGFRi*)2-GAP	Shc	(EGF-EGFRi*)2-GAP-SHC	k22	kd22
(EGF-EGFRi*)2-GAP-SHC	∅	(EGF-EGFRi*)2-GAP-SHC*	k23	kd23
(EGF-EGFRi*)2-GAP-SHC*	Grb2	(EGF-EGFRi*)2-GAP-SHC*-Grb2	k16	kd24
(EGF-EGFRi*)2-GAP-SHC*-Grb2	Sos	(EGF-EGFRi*)2-GAP-SHC*-Grb2-Sos	k25	kd25
(EGF-EGFRi*)2-GAP	Shc*-Grb2-Sos	(EGF-EGFRi*)2-GAP-SHC*-Grb2-Sos	k32	kd32
(EGF-EGFRi*)2-GAP	Grb2-Sos	(EGF-EGFRi*)2-GAP-Grb2-Sos	k34	kd34
(EGF-EGFRi*)2-GAP	Shc*	(EGF-EGFRi*)2-GAP-SHC*	k37	kd37
(EGF-EGFRi*)2-GAP	Shc*-Grb2	(EGF-EGFRi*)2-GAP-SHC*-Grb2	k37	kd37
(EGF-EGFRi*)2-GAP-SHC*	Grb2-Sos	(EGF-EGFRi*)2-GAP-SHC*-Grb2-Sos	k41	kd41
Ras-GTP	GAP	GAP-Ras-GTP	k300	kd20
Ras-GDP	GAP	GAP-Ras-GTP	∅	kd21
Ras-GTPi	GAP	GAP-Ras-GTPi	k300	kd20
Ras-GDP	GAP	GAP-Ras-GTPi	∅	kd21
Ras-GTP*	GAP	GAP-Ras-GTP*	k300	kd20
Ras-GDP	GAP	GAP-Ras-GTP*	∅	kd21
Ras-GTPi*	GAP	GAP-Ras-GTPi*	k300	kd20
Ras-GDP	GAP	GAP-Ras-GTPi*	∅	kd21
Sosi	∅	Sos	k300	∅
phosphatase3mRNA	ERK-PP	ERK-PP	∅	k1000
phosphatase3mRNA	ERKi-PP	ERKi-PP	∅	k1000
∅	∅	phosphatase3	k1001	k1002
phosphatase3mRNA_cyt	phosphatase3	phosphatase3mRNA_cyt	∅	k1003
phosphatase3mRNA	∅	phosphatase3mRNA_cyt	k1004	∅
phosphatase3mRNA	∅	∅	k1005	∅
phosphatase3mRNA_cyt	∅	∅	k1005	∅

Table A.8: Expanded EGFR model Parameters.

Parameter	Value
k0	$0.00 \times 10^{+00}$
kd0	$0.00 \times 10^{+00}$
k1	$3.00 \times 10^{+07}$
kd1	3.84×10^{-03}
k10b	5.43×10^{-02}

Continued on next page

Table A.8 – continued from previous page

Parameter	Value
kd10	1.10×10^{-02}
k2	1.66×10^{-05}
kd2	1.00×10^{-01}
k3	$1.00 \times 10^{+00}$
kd3	1.00×10^{-02}
k4	1.73×10^{-07}
kd4	1.66×10^{-03}
kd5	1.48×10^{-02}
k5	$0.00 \times 10^{+00}$
k6	5.00×10^{-04}
kd6	5.00×10^{-03}
k8	1.66×10^{-06}
kd8	2.00×10^{-01}
k13	$2.17 \times 10^{+00}$
kd13	$0.00 \times 10^{+00}$
k15	$1.00 \times 10^{+04}$
kd15	$0.00 \times 10^{+00}$
k16	1.66×10^{-05}
kd16	$0.00 \times 10^{+00}$
k17	1.66×10^{-05}
kd17	6.00×10^{-02}
k18	2.50×10^{-05}
kd18	$1.30 \times 10^{+00}$
k19	1.66×10^{-07}
kd19	5.00×10^{-01}
k20	3.50×10^{-06}
kd20	4.00×10^{-01}
k21	3.66×10^{-07}
kd21	2.30×10^{-02}
k22	3.50×10^{-05}
kd22	1.00×10^{-01}
k23	$6.00 \times 10^{+00}$
kd23	6.00×10^{-02}
kd24	5.50×10^{-01}
k25	1.66×10^{-05}
kd25	2.14×10^{-02}
k28	1.66×10^{-06}
kd28	5.30×10^{-03}
k29	1.17×10^{-06}
kd29	$1.00 \times 10^{+00}$
kd32	1.00×10^{-01}
k32	4.00×10^{-07}
kd33	2.00×10^{-01}
k33	3.50×10^{-05}
kd34	3.00×10^{-02}
k34	7.50×10^{-06}
kd35	1.50×10^{-03}
k35	7.50×10^{-06}
k36	5.00×10^{-03}
kd36	$0.00 \times 10^{+00}$
kd37	3.00×10^{-01}
k37	1.50×10^{-06}

Continued on next page

Table A.8 – continued from previous page

Parameter	Value
k40	5.00×10^{-05}
kd40	6.40×10^{-02}
k41	5.00×10^{-05}
kd41	4.29×10^{-02}
k42	1.18×10^{-04}
kd42	2.00×10^{-01}
kd43	$1.00 \times 10^{+00}$
k43	$0.00 \times 10^{+00}$
kd44	1.83×10^{-02}
kd45	$3.50 \times 10^{+00}$
k45	$0.00 \times 10^{+00}$
kd47	$2.90 \times 10^{+00}$
k47	$0.00 \times 10^{+00}$
k48	2.38×10^{-05}
kd48	8.00×10^{-01}
kd49	5.80×10^{-02}
k49	$0.00 \times 10^{+00}$
k50	4.50×10^{-07}
kd50	5.00×10^{-01}
kd52	3.30×10^{-02}
kd53	$1.60 \times 10^{+01}$
k53	$0.00 \times 10^{+00}$
kd55	$5.70 \times 10^{+00}$
k55	$0.00 \times 10^{+00}$
kd56	6.00×10^{-01}
k56	2.35×10^{-05}
kd57	2.46×10^{-01}
k57	$0.00 \times 10^{+00}$
k58	8.33×10^{-06}
kd58	5.00×10^{-01}
k52	8.91×10^{-05}
k44	1.96×10^{-05}
k60	5.50×10^{-03}
kd60	$0.00 \times 10^{+00}$
k61	6.70×10^{-04}
kd61	$0.00 \times 10^{+00}$
kd63	2.75×10^{-01}
k63	$0.00 \times 10^{+00}$
k126	1.66×10^{-07}
kd126	$2.00 \times 10^{+00}$
kd127	1.00×10^{-04}
k127	$0.00 \times 10^{+00}$
k200	$2.17 \times 10^{+00}$
kd214	1.81×10^{-04}
kd222	1.97×10^{-04}
kd224	8.25×10^{-05}
kd226	3.01×10^{-05}
kd231	3.00×10^{-05}
kd230	5.43×10^{-05}
k300	1.00×10^{-06}
k1000	$1.00 \times 10^{+00}$
k1001	$1.00 \times 10^{+02}$

Continued on next page

Table A.8 – continued from previous page

Parameter	Value
k1002	1.00×10^{-05}
k1003	1.00×10^{-06}
k1004	1.39×10^{-04}
k1005	4.17×10^{-04}

Table A.9: Expanded EGFR model Species.

Species	Initial Concentration	Is Input
EGF	8.00×10^{-09}	yes
EGFR	$5.00 \times 10^{+04}$	no
EGF-EGFR	$0.00 \times 10^{+00}$	no
(EGF-EGFR)2	$0.00 \times 10^{+00}$	no
(EGF-EGFR*)2	$0.00 \times 10^{+00}$	no
EGFRi	$0.00 \times 10^{+00}$	no
(EGF-EGFR*)2-GAP-Grb2-Prot	$0.00 \times 10^{+00}$	no
(EGF-EGFRi*)2	$0.00 \times 10^{+00}$	no
Proti	$0.00 \times 10^{+00}$	no
EGF-EGFRi	$0.00 \times 10^{+00}$	no
(EGF-EGFRi)2	$0.00 \times 10^{+00}$	no
Prot	$8.10 \times 10^{+04}$	no
GAP	$1.20 \times 10^{+04}$	no
(EGF-EGFR*)2-GAP	$0.00 \times 10^{+00}$	no
EGFi	$0.00 \times 10^{+00}$	no
(EGF-EGFRi*)2-GAP	$0.00 \times 10^{+00}$	no
(EGF-EGFRi*)2-GAP-Grb2	$0.00 \times 10^{+00}$	no
(EGF-EGFRi*)2-GAP-Grb2-Sos	$0.00 \times 10^{+00}$	no
(EGF-EGFRi*)2-GAP-Grb2-Sos-Ras-GDP	$0.00 \times 10^{+00}$	no
(EGF-EGFRi*)2-GAP-Grb2-Sos-Ras-GTP	$0.00 \times 10^{+00}$	no
Grb2	$1.10 \times 10^{+04}$	no
(EGF-EGFR*)2-GAP-Grb2	$0.00 \times 10^{+00}$	no
Sos	$2.63 \times 10^{+04}$	no
(EGF-EGFR*)2-GAP-Grb2-Sos	$0.00 \times 10^{+00}$	no
Ras-GDP	$7.20 \times 10^{+04}$	no
(EGF-EGFR*)2-GAP-Grb2-Sos-Ras-GDP	$0.00 \times 10^{+00}$	no
Ras-GTP	$0.00 \times 10^{+00}$	no
(EGF-EGFR*)2-GAP-Grb2-Sos-Ras-GTP	$0.00 \times 10^{+00}$	no
Grb2-Sos	$4.00 \times 10^{+04}$	no
Shc	$1.01 \times 10^{+05}$	no
(EGF-EGFR*)2-GAP-SHC	$0.00 \times 10^{+00}$	no
(EGF-EGFR*)2-GAP-SHC*	$0.00 \times 10^{+00}$	no
(EGF-EGFR*)2-GAP-SHC*-Grb2	$0.00 \times 10^{+00}$	no
(EGF-EGFR*)2-GAP-SHC*-Grb2-Sos	$0.00 \times 10^{+00}$	no
(EGF-EGFR*)2-GAP-SHC*-Grb2-Sos-Ras-GDP	$0.00 \times 10^{+00}$	no
(EGF-EGFR*)2-GAP-SHC*-Grb2-Sos-Ras-GTP	$0.00 \times 10^{+00}$	no
Shc*-Grb2-Sos	$0.00 \times 10^{+00}$	no
Shc*-Grb2	$0.00 \times 10^{+00}$	no
Shc*	$0.00 \times 10^{+00}$	no
Raf	$4.00 \times 10^{+04}$	no
Raf-Ras-GTP	$0.00 \times 10^{+00}$	no
Ras-GTP*	$0.00 \times 10^{+00}$	no

Continued on next page

Table A.9 – continued from previous page

Species	Initial Concentration	Is Input
Phosphatase1	$4.00 \times 10^{+04}$	no
Raf*	$0.00 \times 10^{+00}$	no
Raf*-phosphatase1	$0.00 \times 10^{+00}$	no
MEK	$2.10 \times 10^{+07}$	no
MEK-Raf*	$0.00 \times 10^{+00}$	no
MEK-P	$0.00 \times 10^{+00}$	no
MEK-P-Raf*	$0.00 \times 10^{+00}$	no
MEK-PP	$0.00 \times 10^{+00}$	no
MEK-PP-phosphatase2	$0.00 \times 10^{+00}$	no
phosphatase2	$4.00 \times 10^{+04}$	no
MEK-P-phosphatase2	$0.00 \times 10^{+00}$	no
ERK	$2.21 \times 10^{+07}$	no
ERK-MEK-PP	$0.00 \times 10^{+00}$	no
ERK-P	$0.00 \times 10^{+00}$	no
ERK-P-MEK-PP	$0.00 \times 10^{+00}$	no
ERK-PP	$0.00 \times 10^{+00}$	no
phosphatase3	$1.00 \times 10^{+07}$	no
ERK-PP-phosphatase3	$0.00 \times 10^{+00}$	no
ERK-P-phosphatase3	$0.00 \times 10^{+00}$	no
(EGF-EGFRi*)2-GAP-SHC	$0.00 \times 10^{+00}$	no
(EGF-EGFRi*)2-GAP-SHC*	$0.00 \times 10^{+00}$	no
(EGF-EGFRi*)2-GAP-SHC*-Grb2	$0.00 \times 10^{+00}$	no
(EGF-EGFRi*)2-GAP-SHC*-Grb2-Sos	$0.00 \times 10^{+00}$	no
(EGF-EGFRi*)2-GAP-SHC*-Grb2-Sos-Ras-GDP	$0.00 \times 10^{+00}$	no
(EGF-EGFRi*)2-GAP-SHC*-Grb2-Sos-Ras-GTP	$0.00 \times 10^{+00}$	no
Ras-GTPi	$0.00 \times 10^{+00}$	no
Raf-Ras-GTPi	$0.00 \times 10^{+00}$	no
Ras-GTPi*	$0.00 \times 10^{+00}$	no
Rafi*	$0.00 \times 10^{+00}$	no
Rafi*-phosphatase1	$0.00 \times 10^{+00}$	no
MEK-Rafi*	$0.00 \times 10^{+00}$	no
MEKi-P	$0.00 \times 10^{+00}$	no
MEK-P-Rafi*	$0.00 \times 10^{+00}$	no
MEKi-PP	$0.00 \times 10^{+00}$	no
MEKi-PP-phosphatase2	$0.00 \times 10^{+00}$	no
MEKi-P-phosphatase2	$0.00 \times 10^{+00}$	no
ERKi-MEKi-PP	$0.00 \times 10^{+00}$	no
ERKi-P	$0.00 \times 10^{+00}$	no
ERKi-P-MEKi-PP	$0.00 \times 10^{+00}$	no
ERKi-PP	$0.00 \times 10^{+00}$	no
ERKi-PP-phosphatase3	$0.00 \times 10^{+00}$	no
ERKi-P-phosphatase3	$0.00 \times 10^{+00}$	no
(EGF-EGFR*)2-GAP-Grb2-Sos-Prot	$0.00 \times 10^{+00}$	no
(EGF-EGFR*)2-GAP-Grb2-Sos-Ras-GDP-Prot	$0.00 \times 10^{+00}$	no
(EGF-EGFR*)2-GAP-Grb2-Sos-Ras-GTP-Prot	$0.00 \times 10^{+00}$	no
(EGF-EGFR*)2-GAP-SHC*-Grb2-Prot	$0.00 \times 10^{+00}$	no
(EGF-EGFR*)2-GAP-SHC*-Grb2-Sos-Prot	$0.00 \times 10^{+00}$	no
(EGF-EGFR*)2-GAP-SHC*-Grb2-Sos-Ras-GDP-Prot	$0.00 \times 10^{+00}$	no
(EGF-EGFR*)2-GAP-SHC*-Grb2-Sos-Ras-GTP-Prot	$0.00 \times 10^{+00}$	no
(EGF-EGFR*)2-GAP-Grb2-Sos-ERK-PP	$0.00 \times 10^{+00}$	no
(EGF-EGFRi*)2-GAP-Grb2-Sos-ERKi-PP	$0.00 \times 10^{+00}$	no
(EGF-EGFR*)2-GAP-SHC*-Grb2-Sos-ERK-PP	$0.00 \times 10^{+00}$	no

Continued on next page

Table A.9 – continued from previous page

Species	Initial Concentration	Is Input
(EGF-EGFRi*)2-GAP-SHC*-Grb2-Sos-ERKi-PP	$0.00 \times 10^{+00}$	no
Sos-ERK-PP	$0.00 \times 10^{+00}$	no
Sos-ERKi-PP	$0.00 \times 10^{+00}$	no
Sosi	$0.00 \times 10^{+00}$	no
GAP-Ras-GTP	$0.00 \times 10^{+00}$	no
GAP-Ras-GTPi	$0.00 \times 10^{+00}$	no
GAP-Ras-GTP*	$0.00 \times 10^{+00}$	no
GAP-Ras-GTPi*	$0.00 \times 10^{+00}$	no
phosphatase3mRNA	$0.00 \times 10^{+00}$	no
phosphatase3mRNA_cyt	$0.00 \times 10^{+00}$	no

A.3 Four Models of a Mitogen Activated Protein Kinase Reaction

Table A.10: MAPK Distributive-Kinase Distributive-Phosphatase Model Reactions.

Reactant 1	Reactant 2	Product	k_{forward}	k_{reverse}
E	S	E:S	k1on	k1off
E	M	E:S	∅	k1cat
E	M	E:M	k2on	k2off
E	D	E:M	∅	k2cat
P	D	P:D	k3on	k3off
P	M	P:D	∅	k2cat
P	M	P:M	k4on	k4off
P	S	P:M	∅	k4cat

Table A.11: MAPK Distributive-Kinase Distributive-Phosphatase Model Parameters

Parameter	Value
k1on	2.00×10^{-02}
k1off	$1.00 \times 10^{+00}$
k1cat	1.00×10^{-02}
k2on	3.20×10^{-02}
k2off	$1.00 \times 10^{+00}$
k2cat	$1.50 \times 10^{+01}$
k3on	4.50×10^{-02}
k3off	$1.00 \times 10^{+00}$
k3cat	9.20×10^{-02}
k4on	1.00×10^{-02}
k4off	$1.00 \times 10^{+00}$
k4cat	5.00×10^{-01}

Table A.12: MAPK (All) Model Species

Species	Initial Concentration	Is Input
E	$1.00 \times 10^{+00}$	yes
S	$2.00 \times 10^{+00}$	no
E:S	$0.00 \times 10^{+00}$	no
M	$0.00 \times 10^{+00}$	no
E:M	$0.00 \times 10^{+00}$	no
D	$0.00 \times 10^{+00}$	no
P	$1.00 \times 10^{+00}$	no
P:D	$0.00 \times 10^{+00}$	no
P:M	$0.00 \times 10^{+00}$	no

A.4 Mass Action Model of *E. coli* Central Central Carbon Metabolism

Table A.13: *E. coli* Model Reactions.

Reactant 1	Reactant 2	Product	k_{forward}	k_{reverse}
F6P	Mur	Mur:F6P	Mur_k0	∅
F6P	Mur	Mur:F6P	∅	Mur_k1
murine	Mur	Mur:F6P	∅	Mur_k2
∅	∅	GAP	Tryp_ksynth	∅
∅	∅	PYR	Tryp_ksynth	∅
∅	∅	PYR	Met_ksynth	∅
RIBU5P	Ru5P	Ru5P:RIBU5P	Ru5P_k0	∅
RIBU5P	Ru5P	Ru5P:RIBU5P	∅	Ru5P_k1
Ru5P:RIBU5P	∅	Ru5P:XYL5P	Ru5P_k2	∅
Ru5P:XYL5P	∅	Ru5P:RIBU5P	Ru5P_k3	∅
XYL5P	Ru5P	Ru5P:XYL5P	∅	Ru5P_k4
XYL5P	Ru5P	Ru5P:XYL5P	Ru5P_k5	∅
RIBU5P	R5P1	R5P1:RIBU5P	R5P1_k0	∅
RIBU5P	R5P1	R5P1:RIBU5P	∅	R5P1_k1
R5P1:RIBU5P	∅	R5P1:RIB5P	R5P1_k2	∅
R5P1:RIB5P	∅	R5P1:RIBU5P	R5P1_k3	∅
RIB5P	R5P1	R5P1:RIB5P	∅	R5P1_k4
RIB5P	R5P1	R5P1:RIB5P	R5P1_k5	∅
RIB5P	TKa	TKa:RIB5P	TKa_k0	∅
RIB5P	TKa	TKa:RIB5P	∅	TKa_k1
TKa:RIB5P	XYL5P	TKa:RIB5P:XYL5P	TKa_k2	∅
TKa:RIB5P	XYL5P	TKa:RIB5P:XYL5P	∅	TKa_k3
TKa:RIB5P:XYL5P	∅	TKa:SED7P:GAP	TKa_k4	∅
TKa:SED7P:GAP	∅	TKa:RIB5P:XYL5P	TKa_k5	∅
TKa:GAP	SED7P	TKa:SED7P:GAP	TKa_k6	∅
TKa:GAP	SED7P	TKa:SED7P:GAP	∅	TKa_k7
GAP	TKa	TKa:GAP	TKa_k8	∅
GAP	TKa	TKa:GAP	∅	TKa_k9
XYL5P	TKb	TKb:XYL5P	TKb_k0	∅
XYL5P	TKb	TKb:XYL5P	∅	TKb_k1

Continued on next page

Table A.13 – continued from previous page

Reactant 1	Reactant 2	Product	k_{forward}	k_{reverse}
TKb:XYL5P	E4P	TKb:XYL5P:E4P	TKb_k2	∅
TKb:XYL5P	E4P	TKb:XYL5P:E4P	∅	TKb_k3
TKb:XYL5P:E4P	∅	TKb:F6P:GAP	TKb_k4	∅
TKb:F6P:GAP	∅	TKb:XYL5P:E4P	TKb_k5	∅
TKb:GAP	F6P	TKb:F6P:GAP	TKb_k6	∅
TKb:GAP	F6P	TKb:F6P:GAP	∅	TKb_k7
GAP	TKb	TKb:GAP	TKb_k8	∅
GAP	TKb	TKb:GAP	∅	TKb_k9
GAP	TA	TA:GAP	TA_k0	∅
GAP	TA	TA:GAP	∅	TA_k1
TA:GAP	SED7P	TA:GAP:SED7P	TA_k2	∅
TA:GAP	SED7P	TA:GAP:SED7P	∅	TA_k3
TA:GAP:SED7P	∅	TA:E4P:F6P	TA_k4	∅
TA:E4P:F6P	∅	TA:GAP:SED7P	TA_k5	∅
TA:F6P	E4P	TA:E4P:F6P	TA_k6	∅
TA:F6P	E4P	TA:E4P:F6P	∅	TA_k7
F6P	TA	TA:F6P	TA_k8	∅
F6P	TA	TA:F6P	∅	TA_k9
RIB5P	RPPK	RPPK:RIB5P	RPPK_k0	∅
RIB5P	RPPK	RPPK:RIB5P	∅	RPPK_k1
nucleotide	RPPK	RPPK:RIB5P	∅	RPPK_k2
G3PDH	DHAP	G3PDH:DHAP	G3PDH_k0	∅
G3PDH	DHAP	G3PDH:DHAP	∅	G3PDH_k1
glycerol	G3PDH	G3PDH:DHAP	∅	G3PDH_k2
SerSynth	PG3	SerSynth:PG3	SerSynth_k0	∅
SerSynth	PG3	SerSynth:PG3	∅	SerSynth_k1
SerSynth	serine	SerSynth:PG3	∅	SerSynth_k2
Synth1	PEP	Synth1:PEP	Synth1_k0	∅
Synth1	PEP	Synth1:PEP	∅	Synth1_k1
cho_mur	Synth1	Synth1:PEP	∅	Synth1_k2
Synth2	PYR	Synth2:PYR	Synth2_k0	∅
Synth2	PYR	Synth2:PYR	∅	Synth2_k1
Synth2	ile	Synth2:PYR	∅	Synth2_k2
G6PDH	G6P	G6PDH:G6P	G6PDH_k0	∅
G6PDH	G6P	G6PDH:G6P	∅	G6PDH_k1
G6PDH	PG	G6PDH:G6P	∅	G6PDH_k2
PDGH	PG	PDGH:PG	PDGH_k0	∅
PDGH	PG	PDGH:PG	∅	PDGH_k1
RIBU5P	PDGH	PDGH:PG	∅	PDGH_k2
DHAP	TIS	TIS:DHAP	TIS_k0	∅
DHAP	TIS	TIS:DHAP	∅	TIS_k1
TIS:DHAP	∅	TIS:GAP	TIS_k2	∅
TIS:GAP	∅	TIS:DHAP	TIS_k3	∅
GAP	TIS	TIS:GAP	∅	TIS_k4
GAP	TIS	TIS:GAP	TIS_k5	∅
PGluMu	PG3	PGluMu:PG3	PGluMu_k0	∅
PGluMu	PG3	PGluMu:PG3	∅	PGluMu_k1
PGluMu:PG3	∅	PGluMu:PG2	PGluMu_k2	∅
PGluMu:PG2	∅	PGluMu:PG3	PGluMu_k3	∅
PGluMu	PG2	PGluMu:PG2	∅	PGluMu_k4
PGluMu	PG2	PGluMu:PG2	PGluMu_k5	∅
PG2	Eno	Eno:PG2	Eno_k0	∅

Continued on next page

Table A.13 – continued from previous page

Reactant 1	Reactant 2	Product	k_{forward}	k_{reverse}
PG2	Eno	Eno:PG2	∅	Eno_k1
Eno:PG2	∅	Eno:PEP	Eno_k2	∅
Eno:PEP	∅	Eno:PG2	Eno_k3	∅
PEP	Eno	Eno:PEP	∅	Eno_k4
PEP	Eno	Eno:PEP	Eno_k5	∅
G6P	PGM	PGM:G6P	PGM_k0	∅
G6P	PGM	PGM:G6P	∅	PGM_k1
PGM:G6P	∅	PGM:G1P	PGM_k2	∅
PGM:G1P	∅	PGM:G6P	PGM_k3	∅
G1P	PGM	PGM:G1P	∅	PGM_k4
G1P	PGM	PGM:G1P	PGM_k5	∅
GAPDH	GAP	GAPDH:GAP	GAPDH_k0	∅
GAPDH	GAP	GAPDH:GAP	∅	GAPDH_k1
GAPDH:GAP	∅	GAPDH:PGP	GAPDH_k2	∅
GAPDH:PGP	∅	GAPDH:GAP	GAPDH_k3	∅
GAPDH	PGP	GAPDH:PGP	∅	GAPDH_k4
GAPDH	PGP	GAPDH:PGP	GAPDH_k5	∅
PGP	PGK	PGK:PGP	PGK_k0	∅
PGP	PGK	PGK:PGP	∅	PGK_k1
PGK:PGP	∅	PGK:PG3	PGK_k2	∅
PGK:PG3	∅	PGK:PGP	PGK_k3	∅
PG3	PGK	PGK:PG3	∅	PGK_k4
PG3	PGK	PGK:PG3	PGK_k5	∅
Aldolase	FDP	Aldolase:FDP	Aldolase_k0	∅
Aldolase	FDP	Aldolase:FDP	∅	Aldolase_k1
Aldolase:FDP	∅	Aldolase:DHAP:GAP	Aldolase_k2	∅
Aldolase:DHAP:GAP	∅	Aldolase:FDP	Aldolase_k3	∅
Aldolase:DHAP	GAP	Aldolase:DHAP:GAP	∅	Aldolase_k4
Aldolase:DHAP	GAP	Aldolase:DHAP:GAP	Aldolase_k5	∅
Aldolase	DHAP	Aldolase:DHAP	∅	Aldolase_k6
Aldolase	DHAP	Aldolase:DHAP	Aldolase_k7	∅
PYR	PDH	PDH:PYR_1	PDH_k0	∅
PYR	PDH	PDH:PYR_1	∅	PDH_k1
PDH:PYR_1	PYR	PDH:PYR_2	PDH_k2	∅
PDH:PYR_1	PYR	PDH:PYR_2	∅	PDH_k3
PDH:PYR_2	PYR	PDH:PYR_3	PDH_k4	∅
PDH:PYR_2	PYR	PDH:PYR_3	∅	PDH_k5
PDH:PYR_3	PYR	PDH:PYR_4	PDH_k6	∅
PDH:PYR_3	PYR	PDH:PYR_4	∅	PDH_k7
accoa	PDH	PDH:PYR_1	∅	PDH_k8
PDH:PYR_1	accoa	PDH:PYR_2	∅	PDH_k9
PDH:PYR_2	accoa	PDH:PYR_3	∅	PDH_k10
PDH:PYR_3	accoa	PDH:PYR_4	∅	PDH_k11
DHAPS00	E4P	DHAPS10	DHAPS_k0	∅
DHAPS00	E4P	DHAPS10	∅	DHAPS_k1
DHAPS00	PEP	DHAPS01	DHAPS_k2	∅
DHAPS00	PEP	DHAPS01	∅	DHAPS_k3
DHAPS01	E4P	DHAPS11	DHAPS_k4	∅
DHAPS01	E4P	DHAPS11	∅	DHAPS_k5
DHAPS01	PEP	DHAPS02	DHAPS_k6	∅
DHAPS01	PEP	DHAPS02	∅	DHAPS_k7
DHAPS02	E4P	DHAPS12	DHAPS_k8	∅

Continued on next page

Table A.13 – continued from previous page

Reactant 1	Reactant 2	Product	k_{forward}	k_{reverse}
DHAPS02	E4P	DHAPS12	∅	DHAPS_k9
DHAPS10	E4P	DHAPS20	DHAPS_k10	∅
DHAPS10	E4P	DHAPS20	∅	DHAPS_k11
DHAPS10	PEP	DHAPS11	DHAPS_k12	∅
DHAPS10	PEP	DHAPS11	∅	DHAPS_k13
DHAPS11	E4P	DHAPS21	DHAPS_k14	∅
DHAPS11	E4P	DHAPS21	∅	DHAPS_k15
DHAPS11	PEP	DHAPS12	DHAPS_k16	∅
DHAPS11	PEP	DHAPS12	∅	DHAPS_k17
DHAPS12	E4P	DHAPS22	DHAPS_k18	∅
DHAPS12	E4P	DHAPS22	∅	DHAPS_k19
DHAPS20	PEP	DHAPS21	DHAPS_k20	∅
DHAPS20	PEP	DHAPS21	∅	DHAPS_k21
DHAPS21	PEP	DHAPS22	DHAPS_k22	∅
DHAPS21	PEP	DHAPS22	∅	DHAPS_k23
DHAPS11	aaa	DHAPS22	∅	DHAPS_k24
PEPCxylase0	FDP	PEPCxylase1	PEPCxylase_k0	∅
PEPCxylase0	FDP	PEPCxylase1	∅	PEPCxylase_k1
PEPCxylase0	PEP	PEPCxylase0A	PEPCxylase_k2	∅
PEPCxylase0	PEP	PEPCxylase0A	∅	PEPCxylase_k3
PEPCxylase0	oaa	PEPCxylase0A	∅	PEPCxylase_k4
PEPCxylase1	FDP	PEPCxylase2	PEPCxylase_k5	∅
PEPCxylase1	FDP	PEPCxylase2	∅	PEPCxylase_k6
PEPCxylase1	PEP	PEPCxylase1A	PEPCxylase_k7	∅
PEPCxylase1	PEP	PEPCxylase1A	∅	PEPCxylase_k8
PEPCxylase1	oaa	PEPCxylase1A	∅	PEPCxylase_k9
PEPCxylase2	FDP	PEPCxylase3	PEPCxylase_k10	∅
PEPCxylase2	FDP	PEPCxylase3	∅	PEPCxylase_k11
PEPCxylase2	PEP	PEPCxylase2A	PEPCxylase_k12	∅
PEPCxylase2	PEP	PEPCxylase2A	∅	PEPCxylase_k13
PEPCxylase2	oaa	PEPCxylase2A	∅	PEPCxylase_k14
PEPCxylase3	FDP	PEPCxylase4	PEPCxylase_k15	∅
PEPCxylase3	FDP	PEPCxylase4	∅	PEPCxylase_k16
PEPCxylase3	PEP	PEPCxylase3A	PEPCxylase_k17	∅
PEPCxylase3	PEP	PEPCxylase3A	∅	PEPCxylase_k18
PEPCxylase3	oaa	PEPCxylase3A	∅	PEPCxylase_k19
PEPCxylase4	PEP	PEPCxylase4A	PEPCxylase_k20	∅
PEPCxylase4	PEP	PEPCxylase4A	∅	PEPCxylase_k21
PEPCxylase4	oaa	PEPCxylase4A	∅	PEPCxylase_k22
G1PAT0	FDP	G1PAT1	G1PAT_k0	∅
G1PAT0	FDP	G1PAT1	∅	G1PAT_k1
G1PAT1	FDP	G1PAT2	G1PAT_k2	∅
G1PAT1	FDP	G1PAT2	∅	G1PAT_k3
G1PAT0	G1P	G1PAT0A	G1PAT_k4	∅
G1PAT0	G1P	G1PAT0A	∅	G1PAT_k5
G1PAT1	G1P	G1PAT1A	G1PAT_k6	∅
G1PAT1	G1P	G1PAT1A	∅	G1PAT_k7
G1PAT2	G1P	G1PAT2A	G1PAT_k8	∅
G1PAT2	G1P	G1PAT2A	∅	G1PAT_k9
PolySac	G1PAT0	G1PAT0A	∅	G1PAT_k10
PolySac	G1PAT1	G1PAT1A	∅	G1PAT_k11
PolySac	G1PAT2	G1PAT2A	∅	G1PAT_k12

Continued on next page

Table A.13 – continued from previous page

Reactant 1	Reactant 2	Product	k_{forward}	k_{reverse}
G6P	PGI	PGI:G6P	PGI_k0	∅
G6P	PGI	PGI:G6P	∅	PGI_k1
F6P	PGI	PGI:G6P	∅	PGI_k2
F6P	PGI	PGI:F6P	PGI_k3	∅
F6P	PGI	PGI:F6P	∅	PGI_k4
G6P	PGI	PGI:F6P	∅	PGI_k5
PGI	PG	PGI:PG	PGI_k6	∅
PGI	PG	PGI:PG	∅	PGI_k7
PGI:PG	PG	PGI:PG:PG	PGI_k8	∅
PGI:PG	PG	PGI:PG:PG	∅	PGI_k9
PGI:PG	F6P	PGI:PG:F6P	PGI_k10	∅
PGI:PG	F6P	PGI:PG:F6P	∅	PGI_k11
PGI:PG	G6P	PGI:PG:F6P	∅	PGI_k12
PGI:PG	G6P	PGI:PG:G6P	PGI_k13	∅
PGI:PG	G6P	PGI:PG:G6P	∅	PGI_k14
PGI:PG	F6P	PGI:PG:G6P	∅	PGI_k15
PFK_0_0	F6P	PFK_1_0	PFK_k0	∅
PFK_1_0	F6P	PFK_2_0	PFK_k1	∅
PFK_2_0	F6P	PFK_3_0	PFK_k2	∅
PFK_3_0	F6P	PFK_4_0	PFK_k3	∅
PFK_4_0	F6P	PFK_5_0	PFK_k4	∅
PFK_5_0	F6P	PFK_6_0	PFK_k5	∅
PFK_6_0	F6P	PFK_7_0	PFK_k6	∅
PFK_7_0	F6P	PFK_8_0	PFK_k7	∅
PFK_8_0	F6P	PFK_9_0	PFK_k8	∅
PFK_9_0	F6P	PFK_10_0	PFK_k9	∅
PFK_10_0	F6P	PFK_11_0	PFK_k10	∅
PFK_0_0	F6P	PFK_1_0	∅	PFK_k11
PFK_1_0	F6P	PFK_2_0	∅	PFK_k12
PFK_2_0	F6P	PFK_3_0	∅	PFK_k13
PFK_3_0	F6P	PFK_4_0	∅	PFK_k14
PFK_4_0	F6P	PFK_5_0	∅	PFK_k15
PFK_5_0	F6P	PFK_6_0	∅	PFK_k16
PFK_6_0	F6P	PFK_7_0	∅	PFK_k17
PFK_7_0	F6P	PFK_8_0	∅	PFK_k18
PFK_8_0	F6P	PFK_9_0	∅	PFK_k19
PFK_9_0	F6P	PFK_10_0	∅	PFK_k20
PFK_10_0	F6P	PFK_11_0	∅	PFK_k21
PFK_0_0	FDP	PFK_1_0	∅	PFK_k22
PFK_1_0	FDP	PFK_2_0	∅	PFK_k23
PFK_2_0	FDP	PFK_3_0	∅	PFK_k24
PFK_3_0	FDP	PFK_4_0	∅	PFK_k25
PFK_4_0	FDP	PFK_5_0	∅	PFK_k26
PFK_5_0	FDP	PFK_6_0	∅	PFK_k27
PFK_6_0	FDP	PFK_7_0	∅	PFK_k28
PFK_7_0	FDP	PFK_8_0	∅	PFK_k29
PFK_8_0	FDP	PFK_9_0	∅	PFK_k30
PFK_9_0	FDP	PFK_10_0	∅	PFK_k31
PFK_10_0	FDP	PFK_11_0	∅	PFK_k32
PFK_0_0	PEP	PFK_0_1	PFK_k33	∅
PFK_0_1	PEP	PFK_0_2	PFK_k34	∅
PFK_0_2	PEP	PFK_0_3	PFK_k35	∅

Continued on next page

Table A.13 – continued from previous page

Reactant 1	Reactant 2	Product	k_{forward}	k_{reverse}
PFK_0_3	PEP	PFK_0_4	PFK_k36	∅
PFK_0_0	PEP	PFK_0_1	∅	PFK_k37
PFK_0_1	PEP	PFK_0_2	∅	PFK_k38
PFK_0_2	PEP	PFK_0_3	∅	PFK_k39
PFK_0_3	PEP	PFK_0_4	∅	PFK_k40
PK00	PEP	PK10	PK_k0	∅
PK10	PEP	PK20	PK_k1	∅
PK20	PEP	PK30	PK_k2	∅
PK30	PEP	PK40	PK_k3	∅
PK01	PEP	PK11	PK_k4	∅
PK11	PEP	PK21	PK_k5	∅
PK21	PEP	PK31	PK_k6	∅
PK31	PEP	PK41	PK_k7	∅
PK00	PEP	PK10	∅	PK_k8
PK10	PEP	PK20	∅	PK_k9
PK20	PEP	PK30	∅	PK_k10
PK30	PEP	PK40	∅	PK_k11
PK01	PEP	PK11	∅	PK_k12
PK11	PEP	PK21	∅	PK_k13
PK21	PEP	PK31	∅	PK_k14
PK31	PEP	PK41	∅	PK_k15
PK00	PYR	PK10	∅	PK_k16
PK10	PYR	PK20	∅	PK_k17
PK20	PYR	PK30	∅	PK_k18
PK30	PYR	PK40	∅	PK_k19
PK01	PYR	PK11	∅	PK_k20
PK11	PYR	PK21	∅	PK_k21
PK21	PYR	PK31	∅	PK_k22
PK31	PYR	PK41	∅	PK_k23
PK00	FDP	PK01	PK_k24	∅
PK00	FDP	PK01	∅	PK_k25
PEP	PTS	PTS:PEP	PTS_k0	∅
PEP	PTS	PTS:PEP	∅	PTS_k1
PTS:PEP	GlcEx	PTS:GlcEx:PEP	PTS_k2	∅
PTS:PEP	GlcEx	PTS:GlcEx:PEP	∅	PTS_k3
PTS:G6P	PYR	PTS:GlcEx:PEP	∅	PTS_k4
G6P	PTS	PTS:G6P	∅	PTS_k5
PYR	PTS	PTS:PYR	PTS_k6	∅
PYR	PTS	PTS:PYR	∅	PTS_k7
G6P	PTS	PTS:I1	PTS_k8	∅
G6P	PTS	PTS:I1	∅	PTS_k9
PTS:I1	G6P	PTS:I2	PTS_k10	∅
PTS:I1	G6P	PTS:I2	∅	PTS_k11
PTS:I2	G6P	PTS:I3	PTS_k12	∅
PTS:I2	G6P	PTS:I3	∅	PTS_k13
PTS:I3	G6P	PTS:I4	PTS_k14	∅
PTS:I3	G6P	PTS:I4	∅	PTS_k15
PTS:PEP	G6P	PTS:PEP:I1	PTS_k16	∅
PTS:PEP	G6P	PTS:PEP:I1	∅	PTS_k17
PTS:PEP:I1	G6P	PTS:PEP:I2	PTS_k18	∅
PTS:PEP:I1	G6P	PTS:PEP:I2	∅	PTS_k19
PTS:PEP:I2	G6P	PTS:PEP:I3	PTS_k20	∅

Continued on next page

Table A.13 – continued from previous page

Reactant 1	Reactant 2	Product	k_{forward}	k_{reverse}
PTS:PEP:I2	G6P	PTS:PEP:I3	∅	PTS_k21
PTS:PEP:I3	G6P	PTS:PEP:I4	PTS_k22	∅
PTS:PEP:I3	G6P	PTS:PEP:I4	∅	PTS_k23
PTS:PYR	G6P	PTS:PYR:I1	PTS_k24	∅
PTS:PYR	G6P	PTS:PYR:I1	∅	PTS_k25
PTS:PYR:I1	G6P	PTS:PYR:I2	PTS_k26	∅
PTS:PYR:I1	G6P	PTS:PYR:I2	∅	PTS_k27
PTS:PYR:I2	G6P	PTS:PYR:I3	PTS_k28	∅
PTS:PYR:I2	G6P	PTS:PYR:I3	∅	PTS_k29
PTS:PYR:I3	G6P	PTS:PYR:I4	PTS_k30	∅
PTS:PYR:I3	G6P	PTS:PYR:I4	∅	PTS_k31
PTS:GlcEx:PEP	G6P	PTS:GlcEx:PEP:I1	PTS_k32	∅
PTS:GlcEx:PEP	G6P	PTS:GlcEx:PEP:I1	∅	PTS_k33
PTS:GlcEx:PEP:I1	G6P	PTS:GlcEx:PEP:I2	PTS_k34	∅
PTS:GlcEx:PEP:I1	G6P	PTS:GlcEx:PEP:I2	∅	PTS_k35
PTS:GlcEx:PEP:I2	G6P	PTS:GlcEx:PEP:I3	PTS_k36	∅
PTS:GlcEx:PEP:I2	G6P	PTS:GlcEx:PEP:I3	∅	PTS_k37
PTS:GlcEx:PEP:I3	G6P	PTS:GlcEx:PEP:I4	PTS_k38	∅
PTS:GlcEx:PEP:I3	G6P	PTS:GlcEx:PEP:I4	∅	PTS_k39
PEP	∅	∅	mu	∅
G6P	∅	∅	mu	∅
PYR	∅	∅	mu	∅
F6P	∅	∅	mu	∅
G1P	∅	∅	mu	∅
PG	∅	∅	mu	∅
FDP	∅	∅	mu	∅
SED7P	∅	∅	mu	∅
GAP	∅	∅	mu	∅
E4P	∅	∅	mu	∅
XYL5P	∅	∅	mu	∅
RIB5P	∅	∅	mu	∅
DHAP	∅	∅	mu	∅
PGP	∅	∅	mu	∅
PG3	∅	∅	mu	∅
PG2	∅	∅	mu	∅
RIBU5P	∅	∅	mu	∅

Table A.14: *E. coli* Model Parameters

Parameter	Value
Aldolase_k0	$1.0000000000000000 \times 10^{e+06}$
Aldolase_k1	$2.1432425069066049 \times 10^{e+11}$
Aldolase_k2	$1.3523049253249044 \times 10^{e+10}$
Aldolase_k3	$3.7515064123159544 \times 10^{e+04}$
Aldolase_k4	$1.3272251897466273 \times 10^{e+04}$
Aldolase_k5	$5.2480865482485516 \times 10^{e+05}$
Aldolase_k6	$5.5142136973924341 \times 10^{e+05}$
Aldolase_k7	$1.5931625876649143 \times 10^{e+05}$
DHAPS_k0	$1.0000000000000000 \times 10^{e+06}$

Continued on next page

Table A.14 – continued from previous page

Parameter	Value
DHAPS_k1	$9.9817107844498411 \times 10^{e+07}$
DHAPS_k10	$1.0000000000000000 \times 10^{e+06}$
DHAPS_k11	$6.6502850754803978 \times 10^{e+02}$
DHAPS_k12	$1.0000000000000000 \times 10^{e+06}$
DHAPS_k13	$2.4520481110314261 \times 10^{e+07}$
DHAPS_k14	$1.0000000000000000 \times 10^{e+06}$
DHAPS_k15	$6.6502850754803978 \times 10^{e+02}$
DHAPS_k16	$1.0000000000000000 \times 10^{e+06}$
DHAPS_k17	$2.3599185370351299 \times 10^{e+02}$
DHAPS_k18	$1.0000000000000000 \times 10^{e+06}$
DHAPS_k19	$6.6502850754803978 \times 10^{e+02}$
DHAPS_k2	$1.0000000000000000 \times 10^{e+06}$
DHAPS_k20	$1.0000000000000000 \times 10^{e+06}$
DHAPS_k21	$2.4520481110314261 \times 10^{e+07}$
DHAPS_k22	$1.0000000000000000 \times 10^{e+06}$
DHAPS_k23	$2.3599185370351299 \times 10^{e+02}$
DHAPS_k24	$1.0795300000000000 \times 10^{e+02}$
DHAPS_k3	$2.4520481110314261 \times 10^{e+07}$
DHAPS_k4	$1.0000000000000000 \times 10^{e+06}$
DHAPS_k5	$9.9817107844498411 \times 10^{e+07}$
DHAPS_k6	$1.0000000000000000 \times 10^{e+06}$
DHAPS_k7	$2.3599185370351299 \times 10^{e+02}$
DHAPS_k8	$1.0000000000000000 \times 10^{e+06}$
DHAPS_k9	$9.9817107844498411 \times 10^{e+07}$
Eno_k0	$5.0000000101437066 \times 10^{e+06}$
Eno_k1	$2.9218301435726549 \times 10^{e+05}$
Eno_k2	$8.0157979662077827 \times 10^{e+05}$
Eno_k3	$3.2094787250456912 \times 10^{e+05}$
Eno_k4	$7.8733175414130231 \times 10^{e+05}$
Eno_k5	$5.0000000093835955 \times 10^{e+06}$
G1PAT_k0	$1.0000000000000000 \times 10^{e+06}$
G1PAT_k1	$2.4304657998068865 \times 10^{e+07}$
G1PAT_k10	$4.9648818237709991 \times 10^{e+00}$
G1PAT_k11	$5.8436405292493612 \times 10^{e+02}$
G1PAT_k12	$5.8436405292493612 \times 10^{e+02}$
G1PAT_k2	$1.0000000000000000 \times 10^{e+06}$
G1PAT_k3	$8.2375643149453239 \times 10^{e+05}$
G1PAT_k4	$1.0000000000000000 \times 10^{e+06}$
G1PAT_k5	$3.2424563390401872 \times 10^{e+06}$
G1PAT_k6	$1.0000000000000000 \times 10^{e+06}$
G1PAT_k7	$3.2424563390401872 \times 10^{e+06}$
G1PAT_k8	$1.0000000000000000 \times 10^{e+06}$
G1PAT_k9	$3.2424563390401872 \times 10^{e+06}$
G3PDH_k0	$1.0000000000000000 \times 10^{e+06}$
G3PDH_k1	$9.9998837960008031 \times 10^{e+05}$
G3PDH_k2	$1.1620400014434582 \times 10^{e+01}$
G6PDH_k0	$1.0000000000000000 \times 10^{e+06}$
G6PDH_k1	$1.4399283648649601 \times 10^{e+07}$
G6PDH_k2	$7.1635135039816817 \times 10^{e+02}$
GAPDH_k0	$1.0000000000006935 \times 10^{e+08}$
GAPDH_k1	$1.4255280428825211 \times 10^{e+10}$
GAPDH_k2	$1.6388988054630449 \times 10^{e+08}$

Continued on next page

Table A.14 – continued from previous page

Parameter	Value
GAPDH_k3	$1.2911085406985128 \times 10^{e+00}$
GAPDH_k4	$7.7999872389368911 \times 10^{e+05}$
GAPDH_k5	$7.5000000000000000 \times 10^{e+10}$
Met_ksynth	$2.2626999999999999 \times 10^{e-03}$
Mur_k0	$1.0000000000000000 \times 10^{e+06}$
Mur_k1	$9.1257800000000007 \times 10^{e+00}$
Mur_k2	$8.7422000000000000 \times 10^{e-01}$
PDGH_k0	$1.0000000000000000 \times 10^{e+06}$
PDGH_k1	$3.7493441681434162 \times 10^{e+07}$
PDGH_k2	$6.6131700196873326 \times 10^{e+03}$
PDH_k0	$1.0000000000000000 \times 10^{e+06}$
PDH_k1	$3.2796807369650561 \times 10^{e+07}$
PDH_k10	$6.1083901753922198 \times 10^{e-03}$
PDH_k11	$6.0591000000000004 \times 10^{e+03}$
PDH_k2	$1.0000000000000000 \times 10^{e+06}$
PDH_k3	$1.3342595718633172 \times 10^{e+08}$
PDH_k4	$1.0000000000000000 \times 10^{e+06}$
PDH_k5	$5.7301676644809306 \times 10^{e+07}$
PDH_k6	$1.0000000000000000 \times 10^{e+06}$
PDH_k7	$1.1604907565945291 \times 10^{e+03}$
PDH_k8	$5.3909705176307341 \times 10^{e+02}$
PDH_k9	$1.3070034036390993 \times 10^{e-01}$
PEPCxylase_k0	$1.0000000000000000 \times 10^{e+06}$
PEPCxylase_k1	$8.0389506995677967 \times 10^{e+09}$
PEPCxylase_k10	$1.0000000000000000 \times 10^{e+06}$
PEPCxylase_k11	$3.7438683203464447 \times 10^{e+04}$
PEPCxylase_k12	$1.0000000000000000 \times 10^{e+06}$
PEPCxylase_k13	$4.2269846243484896 \times 10^{e+06}$
PEPCxylase_k14	$1.5672073331787424 \times 10^{e+04}$
PEPCxylase_k15	$1.0000000000000000 \times 10^{e+06}$
PEPCxylase_k16	$6.1748854911452996 \times 10^{e+04}$
PEPCxylase_k17	$1.0000000000000000 \times 10^{e+06}$
PEPCxylase_k18	$4.2269846243484896 \times 10^{e+06}$
PEPCxylase_k19	$1.5672073331787424 \times 10^{e+04}$
PEPCxylase_k2	$1.0000000000000000 \times 10^{e+06}$
PEPCxylase_k20	$1.0000000000000000 \times 10^{e+06}$
PEPCxylase_k21	$4.2269846243484896 \times 10^{e+06}$
PEPCxylase_k22	$1.5672073331787424 \times 10^{e+04}$
PEPCxylase_k3	$4.2269846243484896 \times 10^{e+06}$
PEPCxylase_k4	$1.0409077486997710 \times 10^{e+02}$
PEPCxylase_k5	$1.0000000000000000 \times 10^{e+06}$
PEPCxylase_k6	$1.8475454057381491 \times 10^{e+06}$
PEPCxylase_k7	$1.0000000000000000 \times 10^{e+06}$
PEPCxylase_k8	$4.2269846243484896 \times 10^{e+06}$
PEPCxylase_k9	$1.5672073331787424 \times 10^{e+04}$
PFK_k0	$1.0000000000000000 \times 10^{e+07}$
PFK_k1	$1.0000000000000000 \times 10^{e+07}$
PFK_k10	$1.0000000000000000 \times 10^{e+07}$
PFK_k11	$2.6712452857997687 \times 10^{e+11}$
PFK_k12	$5.0595719542775881 \times 10^{e+06}$
PFK_k13	$8.7380380875207596 \times 10^{e+06}$
PFK_k14	$1.5472046597139845 \times 10^{e+06}$

Continued on next page

Table A.14 – continued from previous page

Parameter	Value
PFK_k15	$1.0734927311884861 \times 10^{e+06}$
PFK_k16	$1.2802574296156896 \times 10^{e+05}$
PFK_k17	$2.6036228513296116 \times 10^{e+06}$
PFK_k18	$1.2214080272582300 \times 10^{e+07}$
PFK_k19	$7.0377567750365064 \times 10^{e+07}$
PFK_k2	$1.0000000000000000 \times 10^{e+07}$
PFK_k20	$9.9999549409321338 \times 10^{e+11}$
PFK_k21	$9.9999309857040869 \times 10^{e+11}$
PFK_k22	$1.9587802413186693 \times 10^{e+05}$
PFK_k23	$3.9175604826373386 \times 10^{e+05}$
PFK_k24	$5.8763407239560073 \times 10^{e+05}$
PFK_k25	$7.8351209652746771 \times 10^{e+05}$
PFK_k26	$9.7939012065933459 \times 10^{e+05}$
PFK_k27	$1.1752681447912015 \times 10^{e+06}$
PFK_k28	$1.3711461689230686 \times 10^{e+06}$
PFK_k29	$1.5670241930549354 \times 10^{e+06}$
PFK_k3	$1.0000000000000000 \times 10^{e+07}$
PFK_k30	$1.7629022171868025 \times 10^{e+06}$
PFK_k31	$1.9587802413186692 \times 10^{e+06}$
PFK_k32	$2.1546582654505363 \times 10^{e+06}$
PFK_k33	$1.0000000000000000 \times 10^{e+07}$
PFK_k34	$1.0000000000000000 \times 10^{e+07}$
PFK_k35	$1.0000000000000000 \times 10^{e+07}$
PFK_k36	$1.0000000000000000 \times 10^{e+07}$
PFK_k37	$5.9635382316240445 \times 10^{e+06}$
PFK_k38	$1.6965801757957079 \times 10^{e+07}$
PFK_k39	$1.1067662775874595 \times 10^{e+08}$
PFK_k4	$1.0000000000000000 \times 10^{e+07}$
PFK_k40	$3.6151580682876729 \times 10^{e+06}$
PFK_k5	$1.0000000000000000 \times 10^{e+07}$
PFK_k6	$1.0000000000000000 \times 10^{e+07}$
PFK_k7	$1.0000000000000000 \times 10^{e+07}$
PFK_k8	$1.0000000000000000 \times 10^{e+07}$
PFK_k9	$1.0000000000000000 \times 10^{e+07}$
PGI_k0	$1.0000000000000000 \times 10^{e+07}$
PGI_k1	$2.8349012000000000 \times 10^{e+07}$
PGI_k10	$1.0000000000000000 \times 10^{e+07}$
PGI_k11	$9.8243774409480381 \times 10^{e+09}$
PGI_k12	$5.8103971870257437 \times 10^{e+08}$
PGI_k13	$1.0000000000000000 \times 10^{e+07}$
PGI_k14	$5.2308735245413914 \times 10^{e+09}$
PGI_k15	$5.0487078613706931 \times 10^{e+07}$
PGI_k2	$6.5098800000000000 \times 10^{e+05}$
PGI_k3	$1.0000000000000000 \times 10^{e+07}$
PGI_k4	$2.3138474602698651 \times 10^{e+06}$
PGI_k5	$3.4615253973013500 \times 10^{e+05}$
PGI_k6	$1.0000000000000000 \times 10^{e+07}$
PGI_k7	$9.4894223244003777 \times 10^{e+05}$
PGI_k8	$1.0000000000000000 \times 10^{e+07}$
PGI_k9	$4.5388900358918821 \times 10^{e+06}$
PGK_k0	$4.0000000002418436 \times 10^{e+07}$
PGK_k1	$1.1910390732145230 \times 10^{e+06}$

Continued on next page

Table A.14 – continued from previous page

Parameter	Value
PGK_k2	$1.1269927270629425 \times 10^{e+06}$
PGK_k3	$6.7872263734852197 \times 10^{e+04}$
PGK_k4	$4.8336212363022659 \times 10^{e+06}$
PGK_k5	$1.0000000006215300 \times 10^{e+07}$
PGM_k0	$1.0000000000657211 \times 10^{e+06}$
PGM_k1	$1.1057198740453522 \times 10^{e+06}$
PGM_k2	$8.9905333325119841 \times 10^{e+02}$
PGM_k3	$5.6188706748316008 \times 10^{e+01}$
PGM_k4	$1.3544552052271876 \times 10^{e+04}$
PGM_k5	$1.0000000000267429 \times 10^{e+06}$
PGluMu_k0	$5.0000000090073664 \times 10^{e+06}$
PGluMu_k1	$9.1468901122812752 \times 10^{e+05}$
PGluMu_k2	$1.8627073889027213 \times 10^{e+05}$
PGluMu_k3	$2.3599201792045962 \times 10^{e+07}$
PGluMu_k4	$2.1786325514186662 \times 10^{e+07}$
PGluMu_k5	$5.0000000097447606 \times 10^{e+06}$
PK_k0	$1.0000000000000000 \times 10^{e+06}$
PK_k1	$1.0000000000000000 \times 10^{e+06}$
PK_k10	$1.0000000002271533 \times 10^{e+09}$
PK_k11	$1.0225518958698046 \times 10^{e+05}$
PK_k12	$2.9876784545354487 \times 10^{e+05}$
PK_k13	$1.0000000002269371 \times 10^{e+09}$
PK_k14	$1.0000000001866658 \times 10^{e+09}$
PK_k15	$1.0185295983817041 \times 10^{e+05}$
PK_k16	$4.3397280988943159 \times 10^{e+01}$
PK_k17	$4.3397280988943159 \times 10^{e+01}$
PK_k18	$4.3397280988943159 \times 10^{e+01}$
PK_k19	$4.3397280988943159 \times 10^{e+01}$
PK_k2	$1.0000000000000000 \times 10^{e+06}$
PK_k20	$4.3397280988943159 \times 10^{e+01}$
PK_k21	$4.3397280988943159 \times 10^{e+01}$
PK_k22	$4.3397280988943159 \times 10^{e+01}$
PK_k23	$4.3397280988943159 \times 10^{e+01}$
PK_k24	$1.0000000000000000 \times 10^{e+06}$
PK_k25	$9.5936588789919496 \times 10^{e+05}$
PK_k3	$1.0000000000000000 \times 10^{e+06}$
PK_k4	$1.0000000000000000 \times 10^{e+06}$
PK_k5	$1.0000000000000000 \times 10^{e+06}$
PK_k6	$1.0000000000000000 \times 10^{e+06}$
PK_k7	$1.0000000000000000 \times 10^{e+06}$
PK_k8	$7.2423405259541015 \times 10^{e+05}$
PK_k9	$1.0000000001940010 \times 10^{e+09}$
PTS_k0	$1.0000000000000000 \times 10^{e+07}$
PTS_k1	$9.7994562714464098 \times 10^{e+07}$
PTS_k10	$1.0000000000000000 \times 10^{e+07}$
PTS_k11	$4.4431084048532810 \times 10^{e+09}$
PTS_k12	$1.0000000000000000 \times 10^{e+07}$
PTS_k13	$3.5048308261415991 \times 10^{e+06}$
PTS_k14	$1.0000000000000000 \times 10^{e+07}$
PTS_k15	$3.4544908924983547 \times 10^{e+03}$
PTS_k16	$1.0000000000000000 \times 10^{e+07}$
PTS_k17	$7.3663401327365232 \times 10^{e+08}$

Continued on next page

Table A.14 – continued from previous page

Parameter	Value
PTS_k18	$1.0000000000000000 \times 10^{e+07}$
PTS_k19	$4.4431084048532810 \times 10^{e+09}$
PTS_k2	$1.0000000000000000 \times 10^{e+07}$
PTS_k20	$1.0000000000000000 \times 10^{e+07}$
PTS_k21	$3.5048308261415991 \times 10^{e+06}$
PTS_k22	$1.0000000000000000 \times 10^{e+07}$
PTS_k23	$3.4544908924983547 \times 10^{e+03}$
PTS_k24	$1.0000000000000000 \times 10^{e+07}$
PTS_k25	$7.3663401327365232 \times 10^{e+08}$
PTS_k26	$1.0000000000000000 \times 10^{e+07}$
PTS_k27	$4.4431084048532810 \times 10^{e+09}$
PTS_k28	$1.0000000000000000 \times 10^{e+07}$
PTS_k29	$3.5048308261415991 \times 10^{e+06}$
PTS_k3	$9.8206114610983102 \times 10^{e+04}$
PTS_k30	$1.0000000000000000 \times 10^{e+07}$
PTS_k31	$3.4544908924983547 \times 10^{e+03}$
PTS_k32	$1.0000000000000000 \times 10^{e+07}$
PTS_k33	$7.3663401327365232 \times 10^{e+08}$
PTS_k34	$1.0000000000000000 \times 10^{e+07}$
PTS_k35	$4.4431084048532810 \times 10^{e+09}$
PTS_k36	$1.0000000000000000 \times 10^{e+07}$
PTS_k37	$3.5048308261415991 \times 10^{e+06}$
PTS_k38	$1.0000000000000000 \times 10^{e+07}$
PTS_k39	$3.4544908924983547 \times 10^{e+03}$
PTS_k4	$3.8431720438436768 \times 10^{e+17}$
PTS_k5	$1.9216353838636899 \times 10^{e+17}$
PTS_k6	$1.0000000000000000 \times 10^{e+07}$
PTS_k7	$1.8217268567845707 \times 10^{e+06}$
PTS_k8	$1.0000000000000000 \times 10^{e+07}$
PTS_k9	$7.3663401327365232 \times 10^{e+08}$
R5P1_k0	$1.0000000000000000 \times 10^{e+06}$
R5P1_k1	$1.0000000000000000 \times 10^{e+08}$
R5P1_k2	$4.8678508226233616 \times 10^{e+05}$
R5P1_k3	$1.2169627056558404 \times 10^{e+05}$
R5P1_k4	$1.0000000000000000 \times 10^{e+08}$
R5P1_k5	$1.0000000000000000 \times 10^{e+06}$
RPPK_k0	$1.0000000000000000 \times 10^{e+06}$
RPPK_k1	$9.9987099500719385 \times 10^{e+04}$
RPPK_k2	$1.2900500010069370 \times 10^{e+01}$
Ru5P_k0	$1.0000000000000000 \times 10^{e+06}$
Ru5P_k1	$1.0000000000000000 \times 10^{e+08}$
Ru5P_k2	$6.8177933958198258 \times 10^{e+05}$
Ru5P_k3	$4.8698524255855900 \times 10^{e+05}$
Ru5P_k4	$1.0000000000000000 \times 10^{e+08}$
Ru5P_k5	$1.0000000000000000 \times 10^{e+06}$
SerSynth_k0	$1.0000000000000000 \times 10^{e+06}$
SerSynth_k1	$9.9997428790023155 \times 10^{e+05}$
SerSynth_k2	$2.5712100031931431 \times 10^{e+01}$
Synth1_k0	$1.0000000000000000 \times 10^{e+06}$
Synth1_k1	$9.9998046100016276 \times 10^{e+05}$
Synth1_k2	$1.9539000024280771 \times 10^{e+01}$
Synth2_k0	$1.0000000000000000 \times 10^{e+06}$

Continued on next page

Table A.14 – continued from previous page

Parameter	Value
Synth2_k1	$9.9992638140082860 \times 10^{e+05}$
Synth2_k2	$7.3618600090699971 \times 10^{e+01}$
TA_k0	$1.0000000000000000 \times 10^{e+06}$
TA_k1	$3.4778381567798972 \times 10^{e+08}$
TA_k2	$8.9883651167260800 \times 10^{e+06}$
TA_k3	$4.7635749118695432 \times 10^{e+08}$
TA_k4	$3.4703800526606350 \times 10^{e+09}$
TA_k5	$3.7609461023649335 \times 10^{e+09}$
TA_k6	$8.9851047801328618 \times 10^{e+06}$
TA_k7	$4.4990066821897930 \times 10^{e+08}$
TA_k8	$1.0000000000000000 \times 10^{e+06}$
TA_k9	$4.1921580168042737 \times 10^{e+08}$
TIS_k0	$1.0000000041046800 \times 10^{e+06}$
TIS_k1	$3.5595691019993504 \times 10^{e+06}$
TIS_k2	$9.1148290773014291 \times 10^{e+04}$
TIS_k3	$5.4371626067903580 \times 10^{e+03}$
TIS_k4	$2.9514540109794273 \times 10^{e+05}$
TIS_k5	$1.0000000039979001 \times 10^{e+06}$
TKa_k0	$1.0000000000000000 \times 10^{e+06}$
TKa_k1	$2.6036940805652425 \times 10^{e+08}$
TKa_k2	$8.9967360636448544 \times 10^{e+06}$
TKa_k3	$2.8001449621866381 \times 10^{e+08}$
TKa_k4	$1.7756398926550952 \times 10^{e+08}$
TKa_k5	$2.0250717720637748 \times 10^{e+08}$
TKa_k6	$9.1054817665767353 \times 10^{e+06}$
TKa_k7	$3.4125932570363659 \times 10^{e+08}$
TKa_k8	$1.0000000000000000 \times 10^{e+06}$
TKa_k9	$2.9609630557678908 \times 10^{e+08}$
TKb_k0	$1.0000000000000000 \times 10^{e+06}$
TKb_k1	$4.2299426220451057 \times 10^{e+07}$
TKb_k2	$6.1249203330969810 \times 10^{e+06}$
TKb_k3	$1.9782231094656538 \times 10^{e+07}$
TKb_k4	$7.3182174368234470 \times 10^{e+07}$
TKb_k5	$4.7686047850095838 \times 10^{e+07}$
TKb_k6	$6.5235836531243715 \times 10^{e+06}$
TKb_k7	$6.2673145318268642 \times 10^{e+07}$
TKb_k8	$1.0000000000000000 \times 10^{e+06}$
TKb_k9	$9.4166383622396722 \times 10^{e+07}$
Tryp_ksynth	$1.0369999999999999 \times 10^{e-03}$
mu	$2.780000000000000 \times 10^{e-05}$

Table A.15: *E. coli* Model Species

Species	Initial Concentration	Is Input
Aldolase	$0.0000000000000000 \times 10^{e+00}$	no
Aldolase:DHAP	$0.0000000000000000 \times 10^{e+00}$	no
Aldolase:DHAP:GAP	$0.0000000000000000 \times 10^{e+00}$	no
Aldolase:FDP	$0.0000000000000000 \times 10^{e+00}$	no
DHAP	$0.0000000000000000 \times 10^{e+00}$	no

Continued on next page

Table A.15 – continued from previous page

Species	Initial Concentration	Is Input
DHAPS00	$0.0000000000000000 \times 10^{e+00}$	no
DHAPS01	$0.0000000000000000 \times 10^{e+00}$	no
DHAPS02	$0.0000000000000000 \times 10^{e+00}$	no
DHAPS10	$0.0000000000000000 \times 10^{e+00}$	no
DHAPS11	$0.0000000000000000 \times 10^{e+00}$	no
DHAPS12	$0.0000000000000000 \times 10^{e+00}$	no
DHAPS20	$0.0000000000000000 \times 10^{e+00}$	no
DHAPS21	$0.0000000000000000 \times 10^{e+00}$	no
DHAPS22	$0.0000000000000000 \times 10^{e+00}$	no
E4P	$0.0000000000000000 \times 10^{e+00}$	no
Eno	$0.0000000000000000 \times 10^{e+00}$	no
Eno:PEP	$0.0000000000000000 \times 10^{e+00}$	no
Eno:PG2	$0.0000000000000000 \times 10^{e+00}$	no
F6P	$0.0000000000000000 \times 10^{e+00}$	no
FDP	$0.0000000000000000 \times 10^{e+00}$	no
G1P	$0.0000000000000000 \times 10^{e+00}$	no
G1PAT0	$0.0000000000000000 \times 10^{e+00}$	no
G1PAT0A	$0.0000000000000000 \times 10^{e+00}$	no
G1PAT1	$0.0000000000000000 \times 10^{e+00}$	no
G1PAT1A	$0.0000000000000000 \times 10^{e+00}$	no
G1PAT2	$0.0000000000000000 \times 10^{e+00}$	no
G1PAT2A	$0.0000000000000000 \times 10^{e+00}$	no
G3PDH	$0.0000000000000000 \times 10^{e+00}$	no
G3PDH:DHAP	$0.0000000000000000 \times 10^{e+00}$	no
G6P	$0.0000000000000000 \times 10^{e+00}$	no
G6PDH	$0.0000000000000000 \times 10^{e+00}$	no
G6PDH:G6P	$0.0000000000000000 \times 10^{e+00}$	no
GAP	$0.0000000000000000 \times 10^{e+00}$	no
GAPDH	$0.0000000000000000 \times 10^{e+00}$	no
GAPDH:GAP	$0.0000000000000000 \times 10^{e+00}$	no
GAPDH:PGP	$0.0000000000000000 \times 10^{e+00}$	no
GlcEx	$0.0000000000000000 \times 10^{e+00}$	yes
Met	0	no
Mur	$0.0000000000000000 \times 10^{e+00}$	no
Mur:F6P	$0.0000000000000000 \times 10^{e+00}$	no
PDGH	$0.0000000000000000 \times 10^{e+00}$	no
PDGH:PG	$0.0000000000000000 \times 10^{e+00}$	no
PDH	$0.0000000000000000 \times 10^{e+00}$	no
PDH:PYR_1	$0.0000000000000000 \times 10^{e+00}$	no
PDH:PYR_2	$0.0000000000000000 \times 10^{e+00}$	no
PDH:PYR_3	$0.0000000000000000 \times 10^{e+00}$	no
PDH:PYR_4	$0.0000000000000000 \times 10^{e+00}$	no
PEP	$0.0000000000000000 \times 10^{e+00}$	no
PEPCxylase0	$0.0000000000000000 \times 10^{e+00}$	no
PEPCxylase0A	$0.0000000000000000 \times 10^{e+00}$	no
PEPCxylase1	$0.0000000000000000 \times 10^{e+00}$	no
PEPCxylase1A	$0.0000000000000000 \times 10^{e+00}$	no
PEPCxylase2	$0.0000000000000000 \times 10^{e+00}$	no
PEPCxylase2A	$0.0000000000000000 \times 10^{e+00}$	no
PEPCxylase3	$0.0000000000000000 \times 10^{e+00}$	no
PEPCxylase3A	$0.0000000000000000 \times 10^{e+00}$	no
PEPCxylase4	$0.0000000000000000 \times 10^{e+00}$	no

Continued on next page

Table A.15 – continued from previous page

Species	Initial Concentration	Is Input
PEPCxylase4A	$0.0000000000000000 \times 10^{e+00}$	no
PFK_0_0	$0.0000000000000000 \times 10^{e+00}$	no
PFK_0_1	$0.0000000000000000 \times 10^{e+00}$	no
PFK_0_2	$0.0000000000000000 \times 10^{e+00}$	no
PFK_0_3	$0.0000000000000000 \times 10^{e+00}$	no
PFK_0_4	$0.0000000000000000 \times 10^{e+00}$	no
PFK_10_0	$0.0000000000000000 \times 10^{e+00}$	no
PFK_11_0	$0.0000000000000000 \times 10^{e+00}$	no
PFK_1_0	$0.0000000000000000 \times 10^{e+00}$	no
PFK_2_0	$0.0000000000000000 \times 10^{e+00}$	no
PFK_3_0	$0.0000000000000000 \times 10^{e+00}$	no
PFK_4_0	$0.0000000000000000 \times 10^{e+00}$	no
PFK_5_0	$0.0000000000000000 \times 10^{e+00}$	no
PFK_6_0	$0.0000000000000000 \times 10^{e+00}$	no
PFK_7_0	$0.0000000000000000 \times 10^{e+00}$	no
PFK_8_0	$0.0000000000000000 \times 10^{e+00}$	no
PFK_9_0	$0.0000000000000000 \times 10^{e+00}$	no
PG	$0.0000000000000000 \times 10^{e+00}$	no
PG2	$0.0000000000000000 \times 10^{e+00}$	no
PG3	$0.0000000000000000 \times 10^{e+00}$	no
PGI	$0.0000000000000000 \times 10^{e+00}$	no
PGI:F6P	$0.0000000000000000 \times 10^{e+00}$	no
PGI:G6P	$0.0000000000000000 \times 10^{e+00}$	no
PGI:PG	$0.0000000000000000 \times 10^{e+00}$	no
PGI:PG:F6P	$0.0000000000000000 \times 10^{e+00}$	no
PGI:PG:G6P	$0.0000000000000000 \times 10^{e+00}$	no
PGI:PG:PG	$0.0000000000000000 \times 10^{e+00}$	no
PGK	$0.0000000000000000 \times 10^{e+00}$	no
PGK:PG3	$0.0000000000000000 \times 10^{e+00}$	no
PGK:PGP	$0.0000000000000000 \times 10^{e+00}$	no
PGM	$0.0000000000000000 \times 10^{e+00}$	no
PGM:G1P	$0.0000000000000000 \times 10^{e+00}$	no
PGM:G6P	$0.0000000000000000 \times 10^{e+00}$	no
PGP	$0.0000000000000000 \times 10^{e+00}$	no
PGluMu	$0.0000000000000000 \times 10^{e+00}$	no
PGluMu:PG2	$0.0000000000000000 \times 10^{e+00}$	no
PGluMu:PG3	$0.0000000000000000 \times 10^{e+00}$	no
PK00	$0.0000000000000000 \times 10^{e+00}$	no
PK01	$0.0000000000000000 \times 10^{e+00}$	no
PK10	$0.0000000000000000 \times 10^{e+00}$	no
PK11	$0.0000000000000000 \times 10^{e+00}$	no
PK20	$0.0000000000000000 \times 10^{e+00}$	no
PK21	$0.0000000000000000 \times 10^{e+00}$	no
PK30	$0.0000000000000000 \times 10^{e+00}$	no
PK31	$0.0000000000000000 \times 10^{e+00}$	no
PK40	$0.0000000000000000 \times 10^{e+00}$	no
PK41	$0.0000000000000000 \times 10^{e+00}$	no
PTS	$0.0000000000000000 \times 10^{e+00}$	no
PTS:G6P	$0.0000000000000000 \times 10^{e+00}$	no
PTS:GlcEx:PEP	$0.0000000000000000 \times 10^{e+00}$	no
PTS:GlcEx:PEP:I1	$0.0000000000000000 \times 10^{e+00}$	no
PTS:GlcEx:PEP:I2	$0.0000000000000000 \times 10^{e+00}$	no

Continued on next page

Table A.15 – continued from previous page

Species	Initial Concentration	Is Input
PTS:GlcEx:PEP:I3	$0.0000000000000000 \times 10^{e+00}$	no
PTS:GlcEx:PEP:I4	$0.0000000000000000 \times 10^{e+00}$	no
PTS:I1	$0.0000000000000000 \times 10^{e+00}$	no
PTS:I2	$0.0000000000000000 \times 10^{e+00}$	no
PTS:I3	$0.0000000000000000 \times 10^{e+00}$	no
PTS:I4	$0.0000000000000000 \times 10^{e+00}$	no
PTS:PEP	$0.0000000000000000 \times 10^{e+00}$	no
PTS:PEP:I1	$0.0000000000000000 \times 10^{e+00}$	no
PTS:PEP:I2	$0.0000000000000000 \times 10^{e+00}$	no
PTS:PEP:I3	$0.0000000000000000 \times 10^{e+00}$	no
PTS:PEP:I4	$0.0000000000000000 \times 10^{e+00}$	no
PTS:PYR	$0.0000000000000000 \times 10^{e+00}$	no
PTS:PYR:I1	$0.0000000000000000 \times 10^{e+00}$	no
PTS:PYR:I2	$0.0000000000000000 \times 10^{e+00}$	no
PTS:PYR:I3	$0.0000000000000000 \times 10^{e+00}$	no
PTS:PYR:I4	$0.0000000000000000 \times 10^{e+00}$	no
PYR	$0.0000000000000000 \times 10^{e+00}$	no
PolySac	$0.0000000000000000 \times 10^{e+00}$	no
R5P1	$0.0000000000000000 \times 10^{e+00}$	no
R5P1:RIB5P	$0.0000000000000000 \times 10^{e+00}$	no
R5P1:RIBU5P	$0.0000000000000000 \times 10^{e+00}$	no
RIB5P	$0.0000000000000000 \times 10^{e+00}$	no
RIBU5P	$0.0000000000000000 \times 10^{e+00}$	no
RPPK	$0.0000000000000000 \times 10^{e+00}$	no
RPPK:RIB5P	$0.0000000000000000 \times 10^{e+00}$	no
Ru5P	$0.0000000000000000 \times 10^{e+00}$	no
Ru5P:RIBU5P	$0.0000000000000000 \times 10^{e+00}$	no
Ru5P:XYL5P	$0.0000000000000000 \times 10^{e+00}$	no
SED7P	$0.0000000000000000 \times 10^{e+00}$	no
SerSynth	$0.0000000000000000 \times 10^{e+00}$	no
SerSynth:PG3	$0.0000000000000000 \times 10^{e+00}$	no
Synth1	$0.0000000000000000 \times 10^{e+00}$	no
Synth1:PEP	$0.0000000000000000 \times 10^{e+00}$	no
Synth2	$0.0000000000000000 \times 10^{e+00}$	no
Synth2:PYR	$0.0000000000000000 \times 10^{e+00}$	no
TA	$0.0000000000000000 \times 10^{e+00}$	no
TA:E4P:F6P	$0.0000000000000000 \times 10^{e+00}$	no
TA:F6P	$0.0000000000000000 \times 10^{e+00}$	no
TA:GAP	$0.0000000000000000 \times 10^{e+00}$	no
TA:GAP:SED7P	$0.0000000000000000 \times 10^{e+00}$	no
TIS	$0.0000000000000000 \times 10^{e+00}$	no
TIS:DHAP	$0.0000000000000000 \times 10^{e+00}$	no
TIS:GAP	$0.0000000000000000 \times 10^{e+00}$	no
TKa	$0.0000000000000000 \times 10^{e+00}$	no
TKa:GAP	$0.0000000000000000 \times 10^{e+00}$	no
TKa:RIB5P	$0.0000000000000000 \times 10^{e+00}$	no
TKa:RIB5P:XYL5P	$0.0000000000000000 \times 10^{e+00}$	no
TKa:SED7P:GAP	$0.0000000000000000 \times 10^{e+00}$	no
TKb	$0.0000000000000000 \times 10^{e+00}$	no
TKb:F6P:GAP	$0.0000000000000000 \times 10^{e+00}$	no
TKb:GAP	$0.0000000000000000 \times 10^{e+00}$	no
TKb:XYL5P	$0.0000000000000000 \times 10^{e+00}$	no

Continued on next page

Table A.15 – continued from previous page

Species	Initial Concentration	Is Input
TKb:XYL5P:E4P	$0.0000000000000000 \times 10^{e+00}$	no
Tryp	0	no
XYL5P	$0.0000000000000000 \times 10^{e+00}$	no
aaa	$0.0000000000000000 \times 10^{e+00}$	no
accoa	$0.0000000000000000 \times 10^{e+00}$	no
cho_mur	$0.0000000000000000 \times 10^{e+00}$	no
glycerol	$0.0000000000000000 \times 10^{e+00}$	no
ile	$0.0000000000000000 \times 10^{e+00}$	no
murine	$0.0000000000000000 \times 10^{e+00}$	no
nucleotide	$0.0000000000000000 \times 10^{e+00}$	no
oaa	$0.0000000000000000 \times 10^{e+00}$	no
serine	$0.0000000000000000 \times 10^{e+00}$	no

Appendix B

Mass Action Model of *Escherichia coli* Central Carbon Metabolism: Enzyme Fits

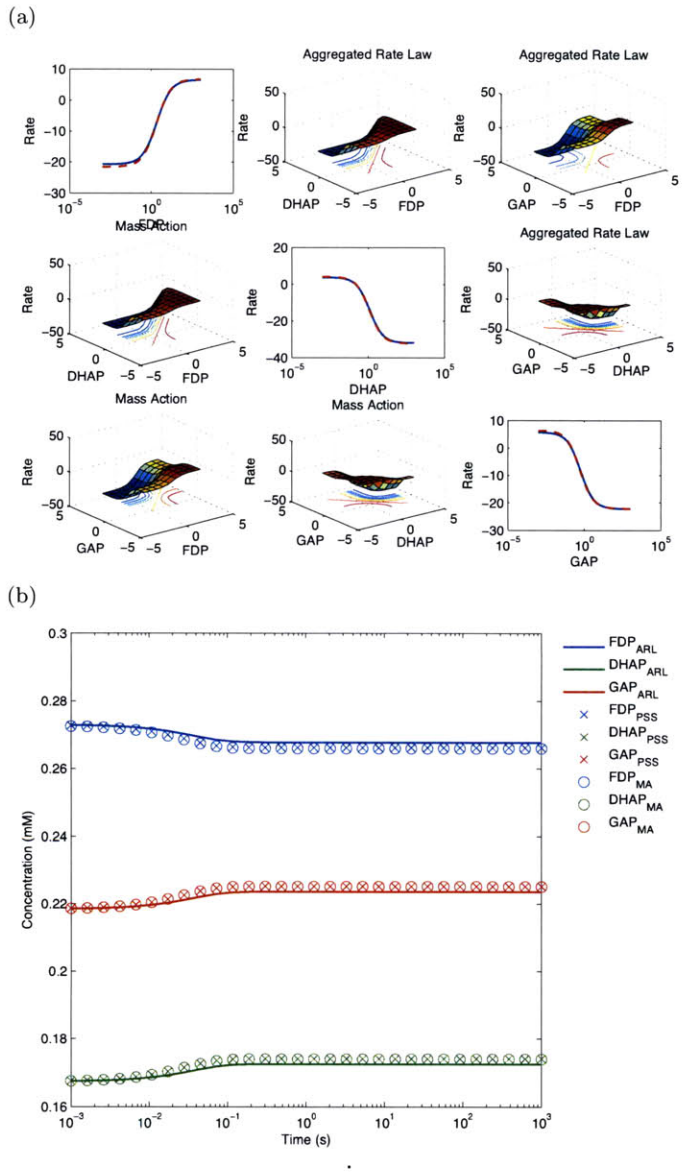


Figure B-1: Fit of Aldolase Mass Action Mechanism to Aggregated Rate Law Mechanism. Rate vs substrate plot (a), and simulation from initial conditions (b)

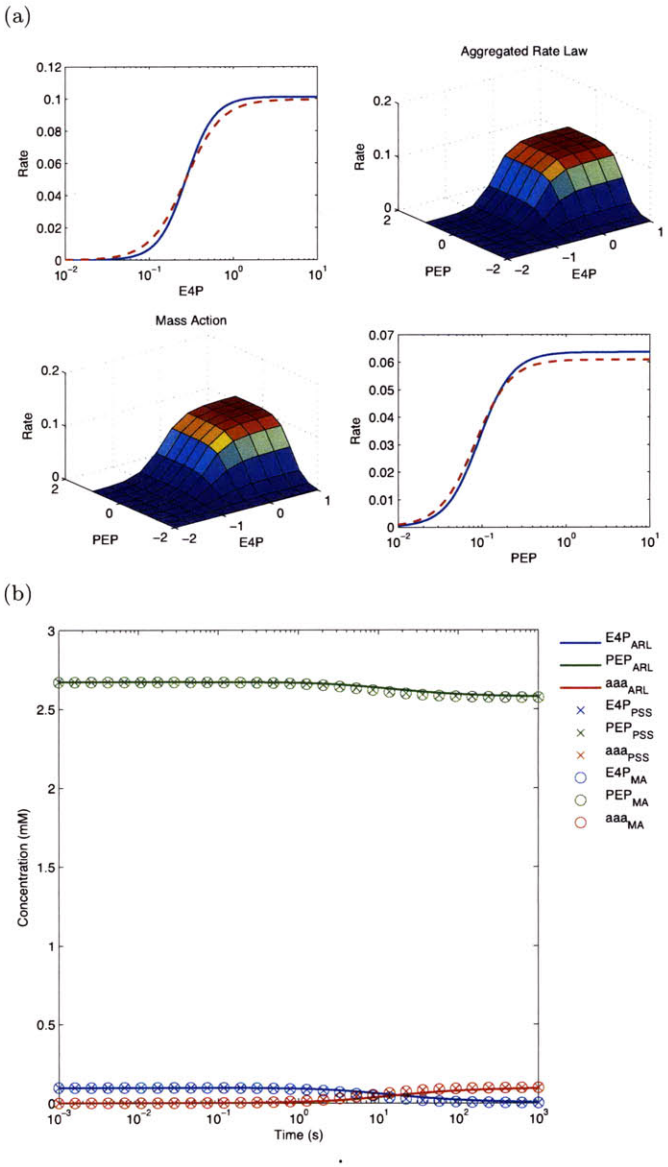


Figure B-2: Fit of DAHAPS Mass Action Mechanism to Aggregated Rate Law Mechanism. Rate vs substrate plot (a), and simulation from initial conditions (b)

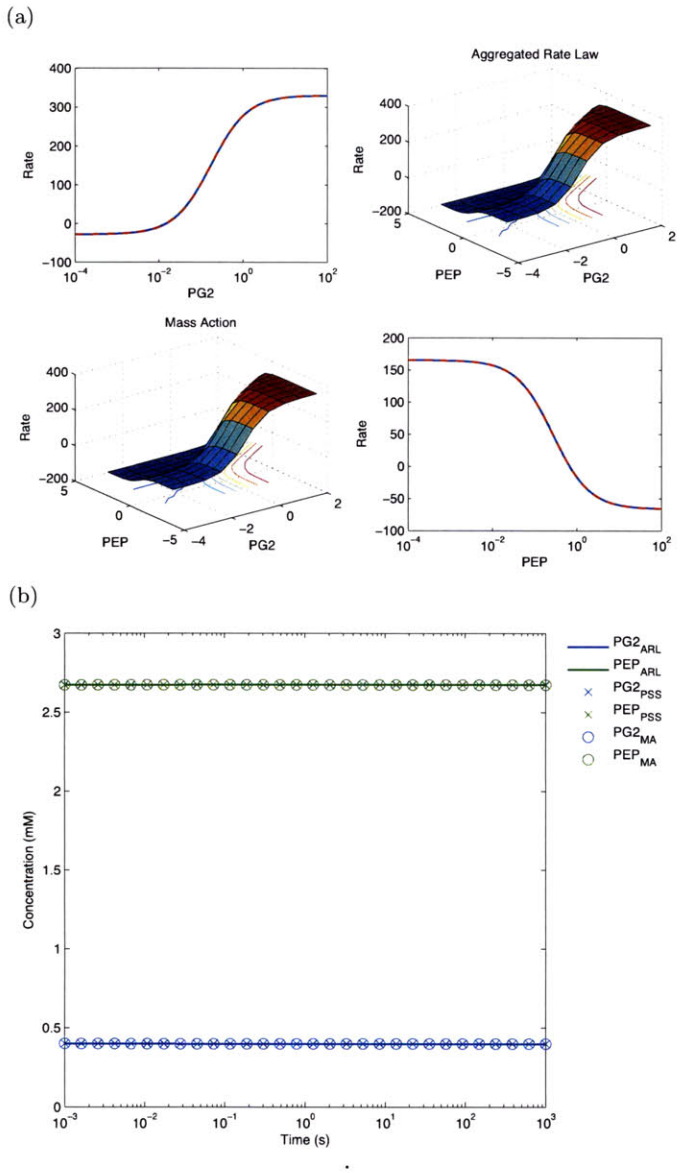


Figure B-3: Fit of Eno Mass Action Mechanism to Aggregated Rate Law Mechanism. Rate vs substrate plot (a), and simulation from initial conditions (b)

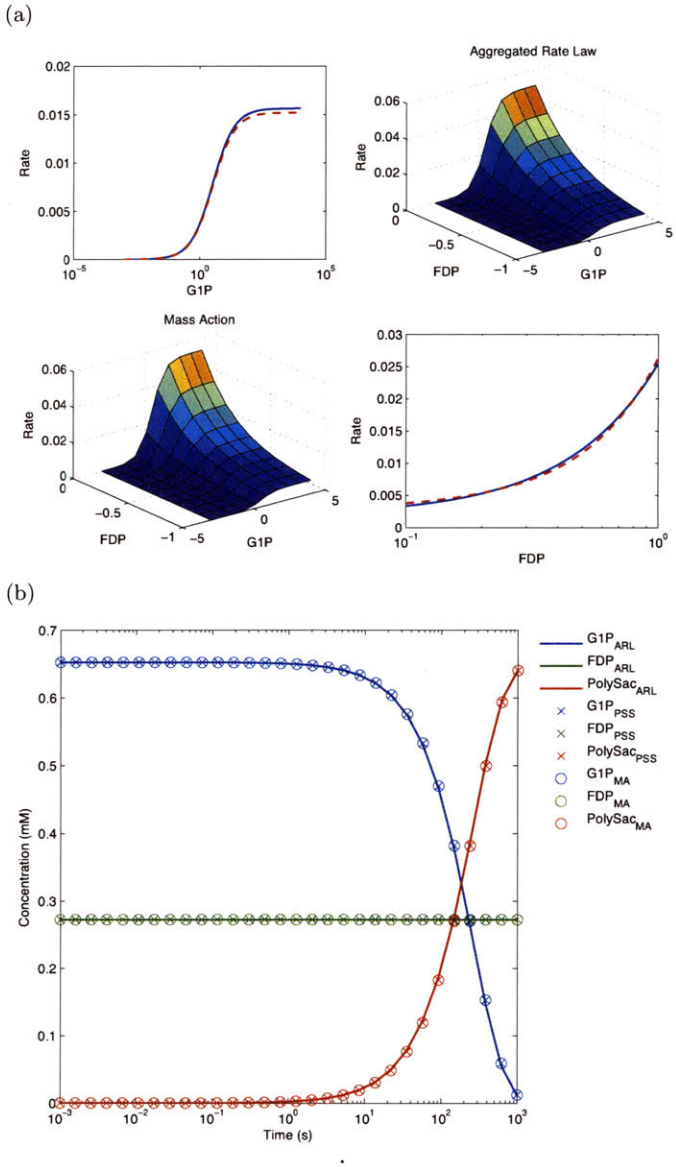


Figure B-4: Fit of G1PAT Mass Action Mechanism to Aggregated Rate Law Mechanism. Rate vs substrate plot (a), and simulation from initial conditions (b)

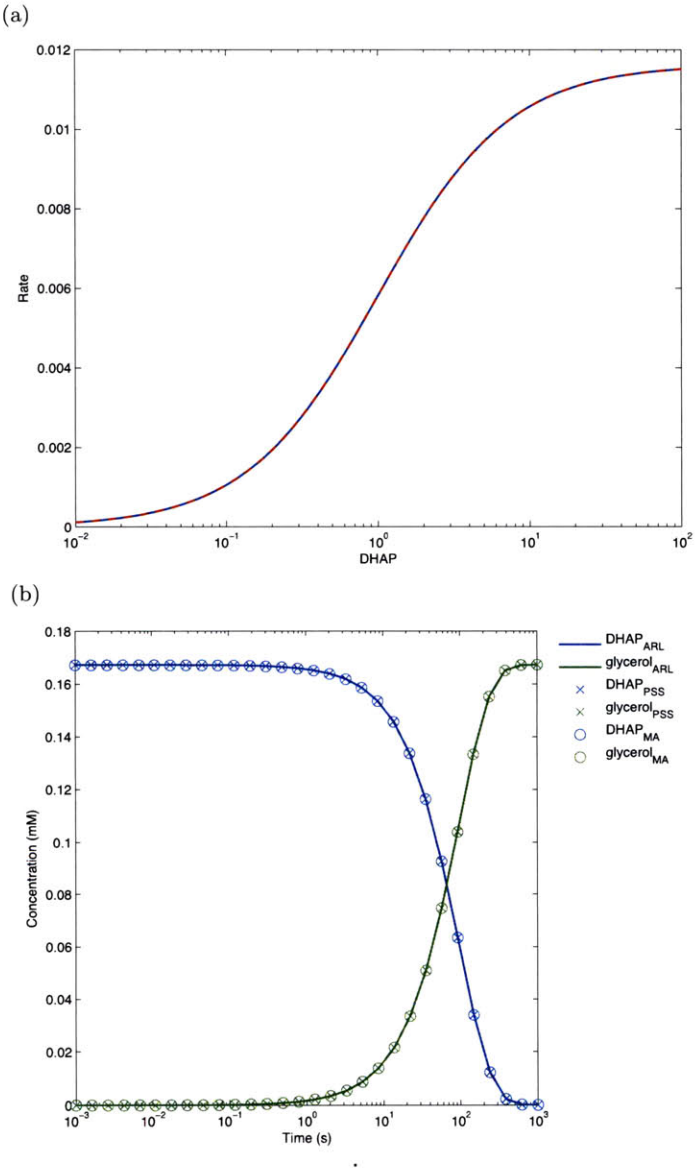


Figure B-5: Fit of G3PDH Mass Action Mechanism to Aggregated Rate Law Mechanism. Rate vs substrate plot (a), and simulation from initial conditions (b)

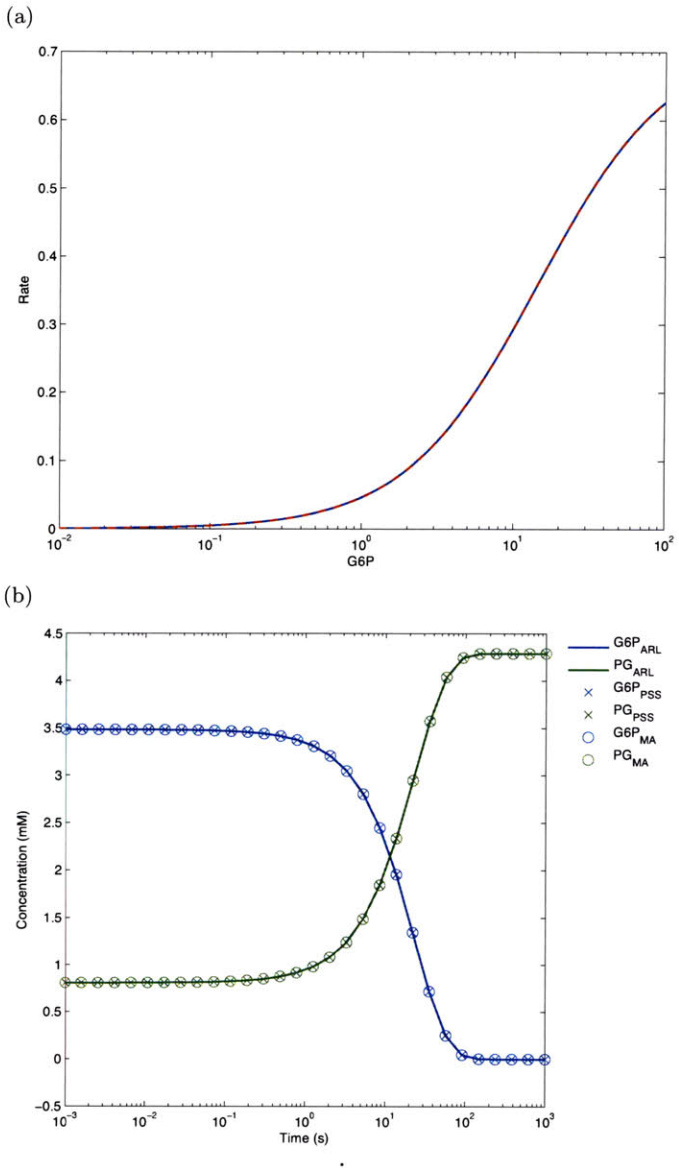


Figure B-6: Fit of G6PDH Mass Action Mechanism to Aggregated Rate Law Mechanism. Rate vs substrate plot (a), and simulation from initial conditions (b)

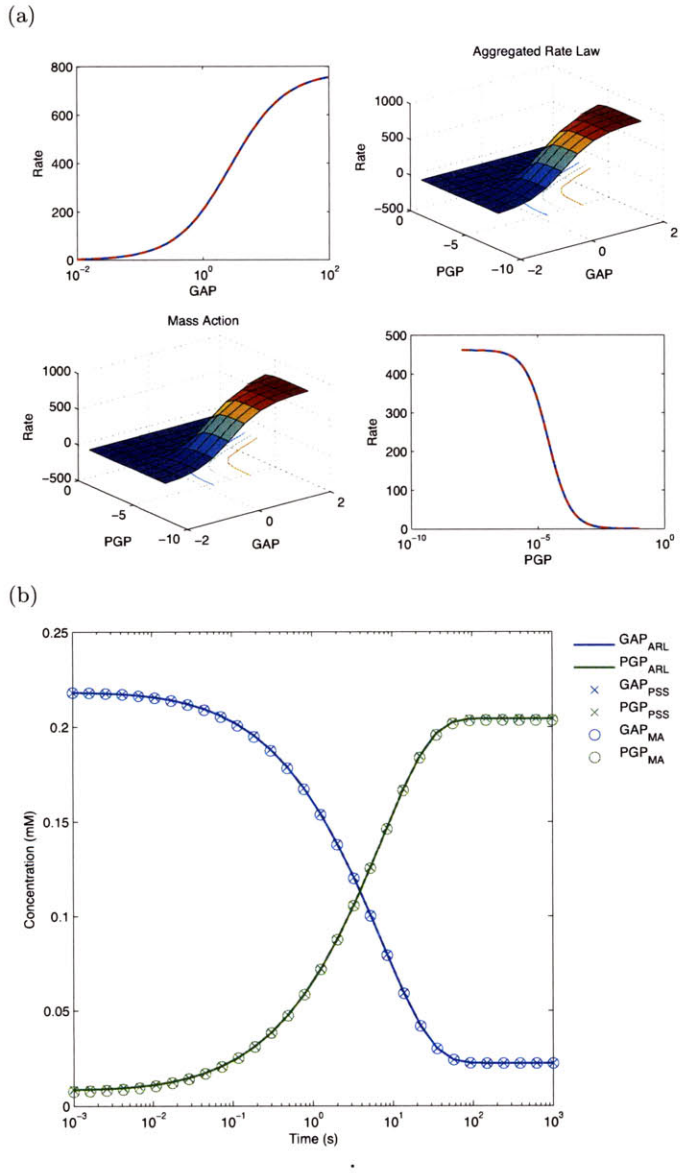


Figure B-7: Fit of GADPH Mass Action Mechanism to Aggregated Rate Law Mechanism. Rate vs substrate plot (a), and simulation from initial conditions (b)

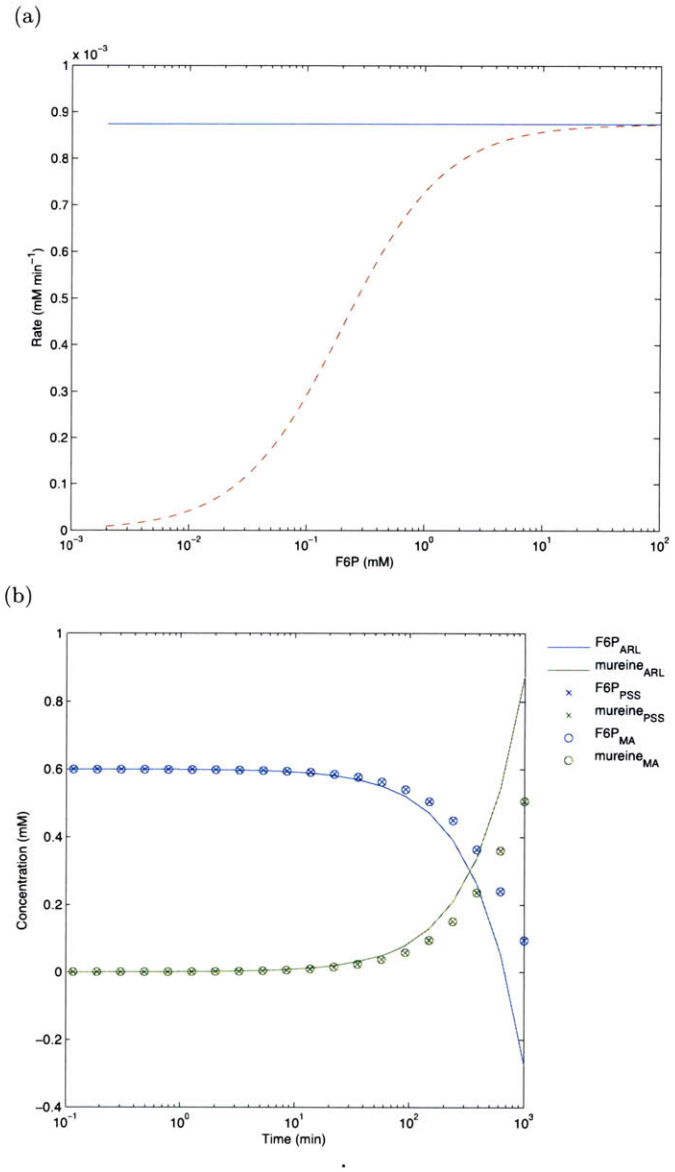


Figure B-8: Fit of Mur Mass Action Mechanism to Aggregated Rate Law Mechanism. Rate vs substrate plot (a), and simulation from initial conditions (b)

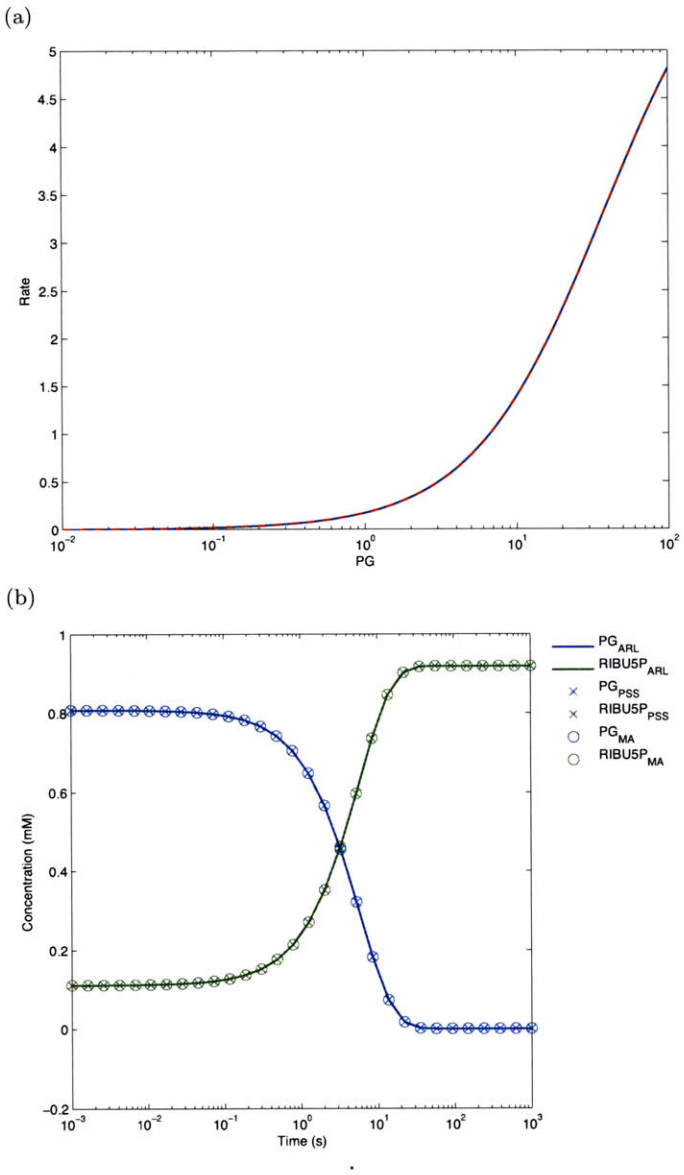


Figure B-9: Fit of PDGH Mass Action Mechanism to Aggregated Rate Law Mechanism. Rate vs substrate plot (a), and simulation from initial conditions (b)

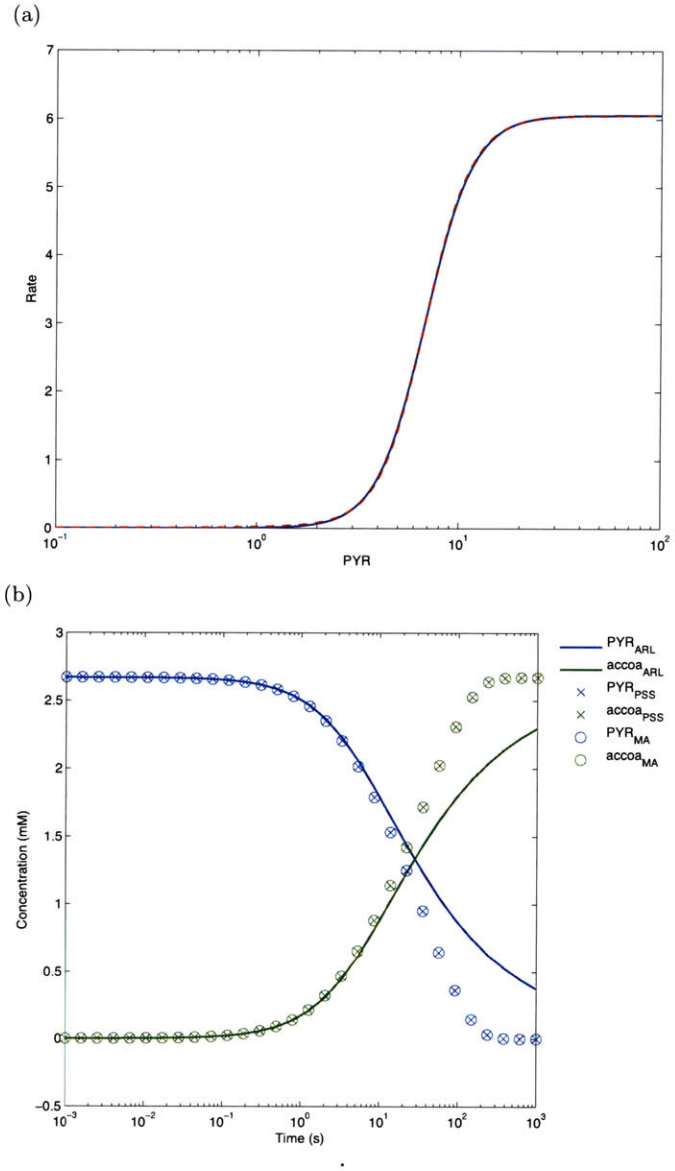


Figure B-10: Fit of PDH Mass Action Mechanism to Aggregated Rate Law Mechanism. Rate vs substrate plot (a), and simulation from initial conditions (b)

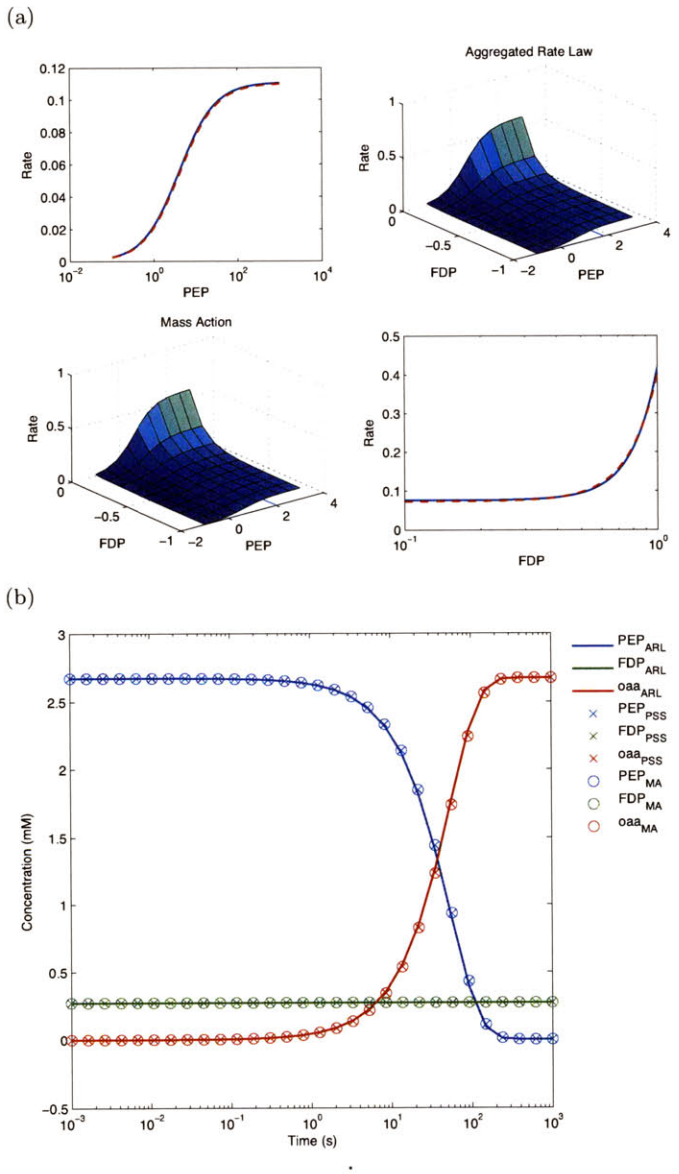


Figure B-11: Fit of PEPCxylase Mass Action Mechanism to Aggregated Rate Law Mechanism. Rate vs substrate plot (a), and simulation from initial conditions (b)

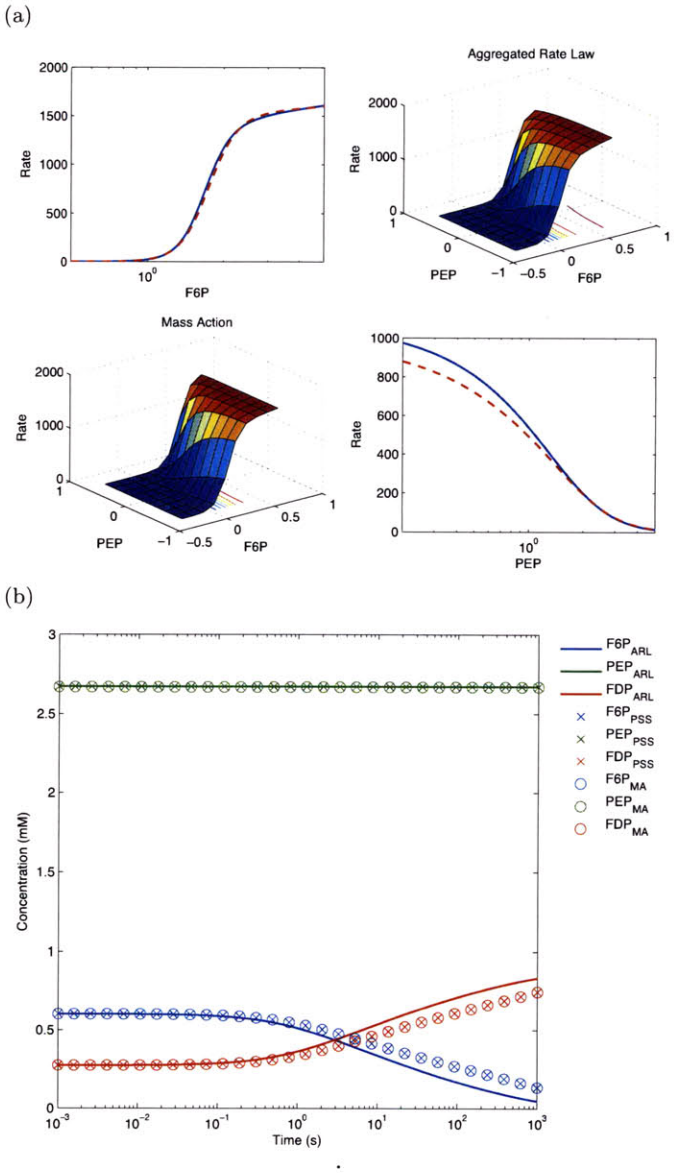


Figure B-12: Fit of PFK Mass Action Mechanism to Aggregated Rate Law Mechanism. Rate vs substrate plot (a), and simulation from initial conditions (b)

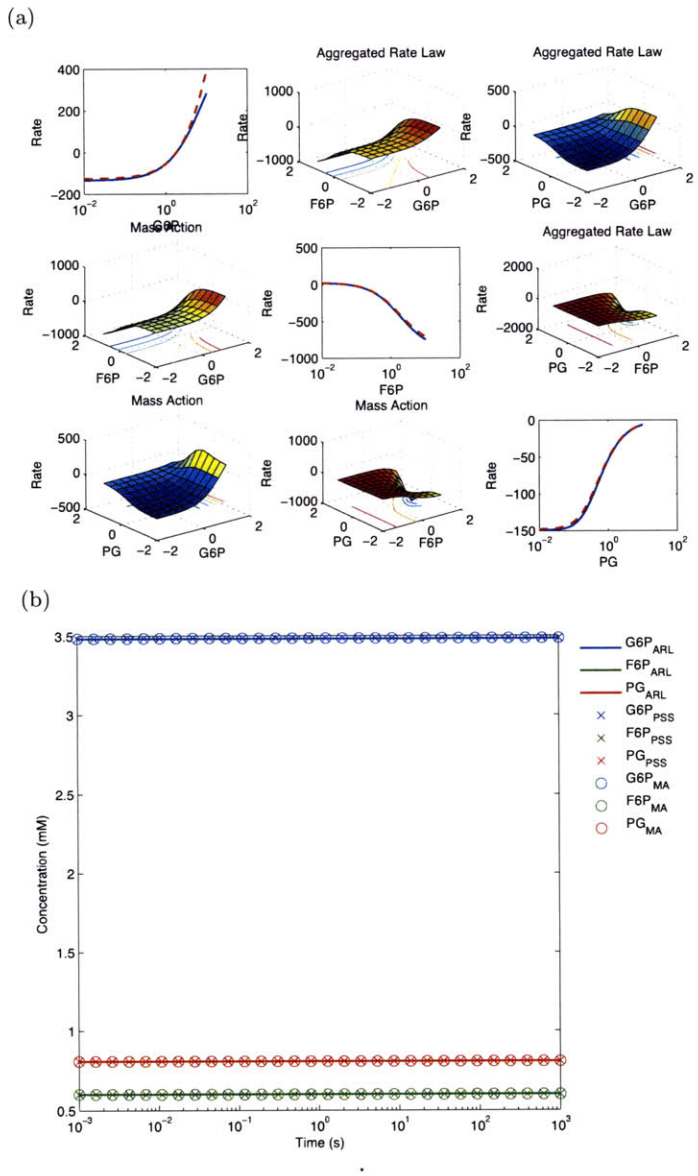


Figure B-13: Fit of PGI Mass Action Mechanism to Aggregated Rate Law Mechanism. Rate vs substrate plot (a), and simulation from initial conditions (b)

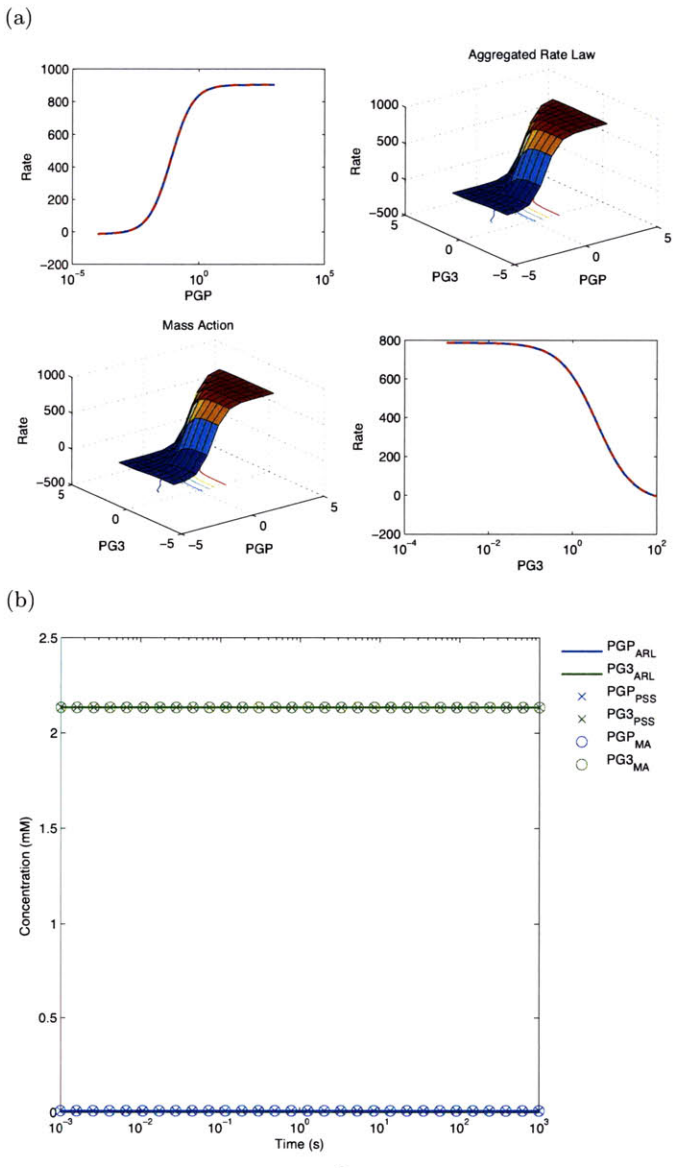


Figure B-14: Fit of PGK Mass Action Mechanism to Aggregated Rate Law Mechanism. Rate vs substrate plot (a), and simulation from initial conditions (b)

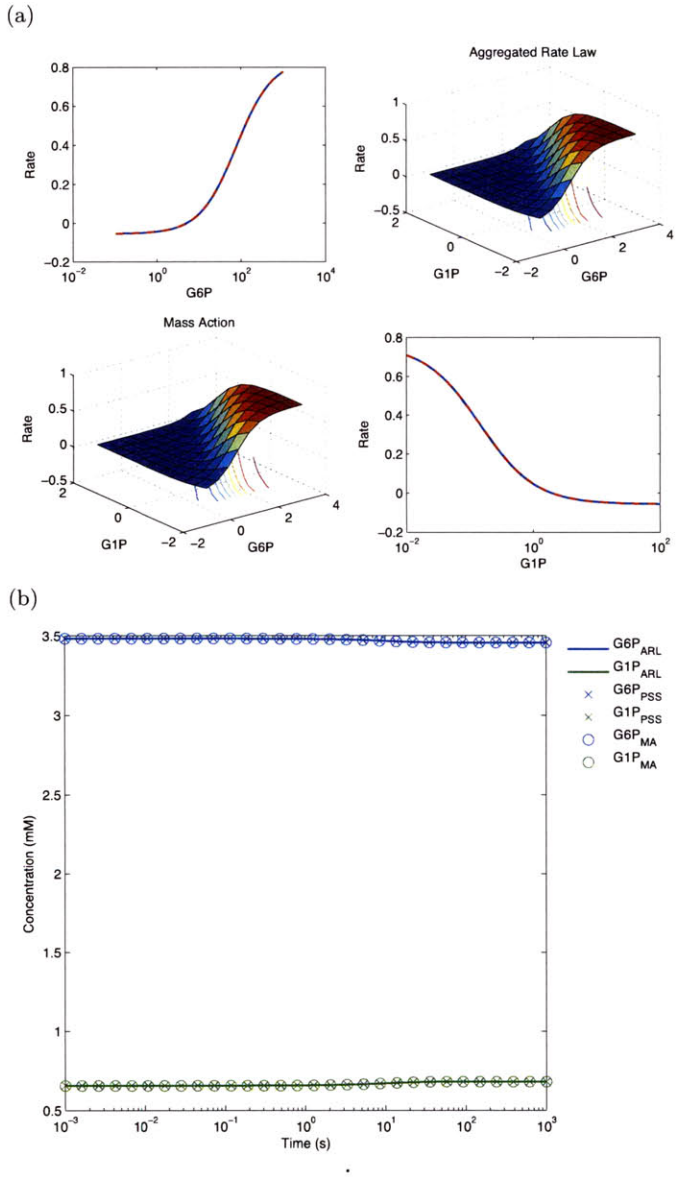


Figure B-15: Fit of PGM Mass Action Mechanism to Aggregated Rate Law Mechanism. Rate vs substrate plot (a), and simulation from initial conditions (b)

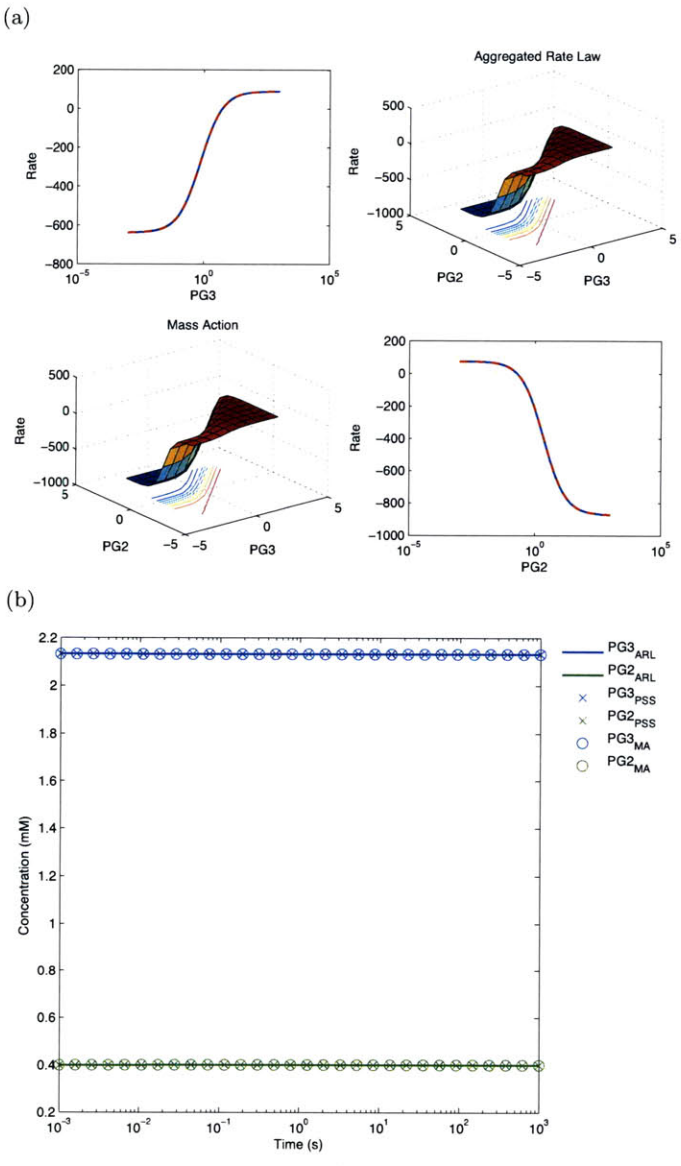


Figure B-16: Fit of PGluMu Mass Action Mechanism to Aggregated Rate Law Mechanism. Rate vs substrate plot (a), and simulation from initial conditions (b)

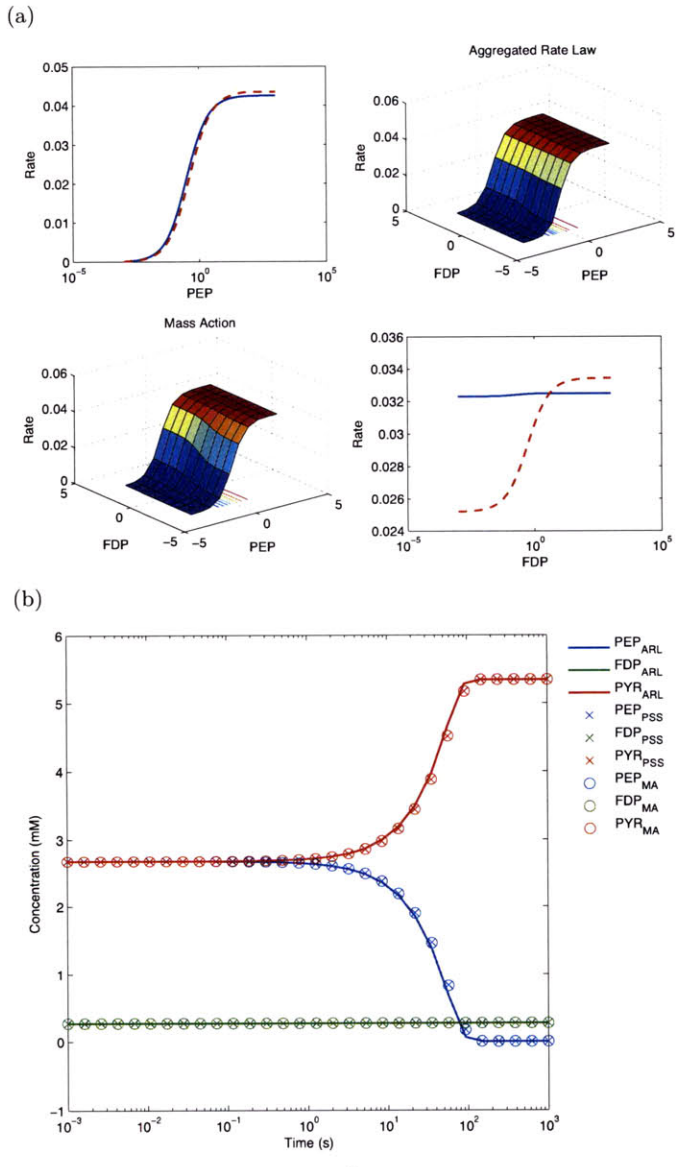


Figure B-17: Fit of PK Mass Action Mechanism to Aggregated Rate Law Mechanism. Rate vs substrate plot (a), and simulation from initial conditions (b)

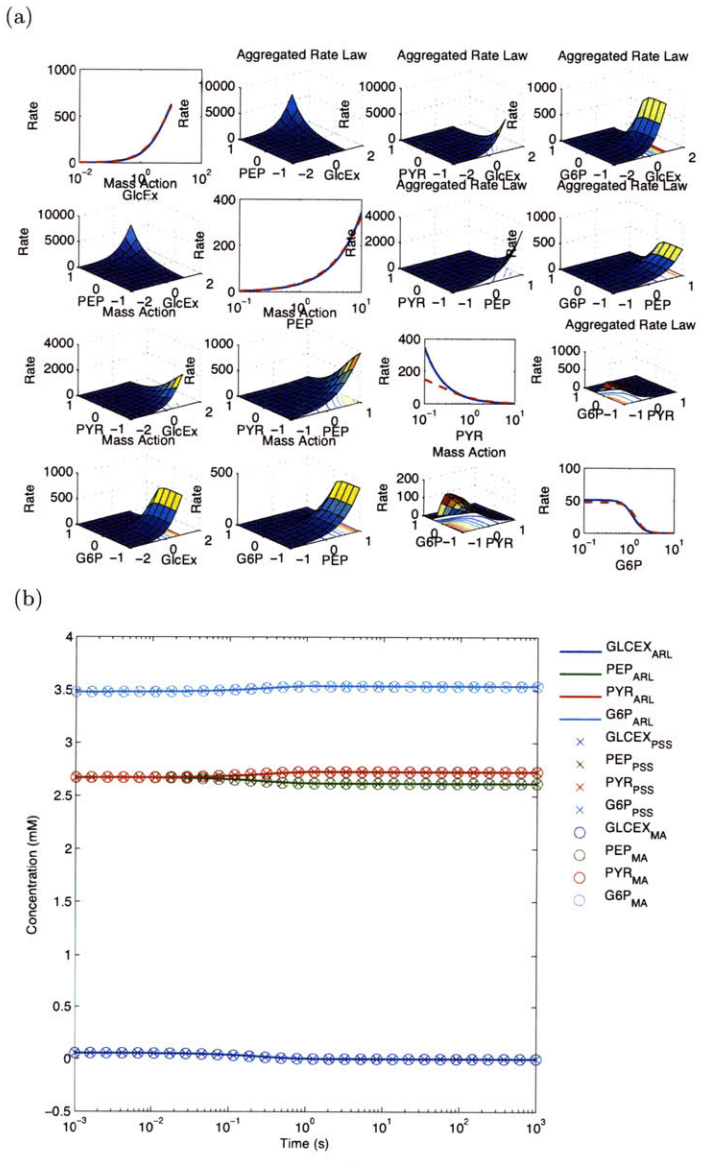


Figure B-18: Fit of PTS Mass Action Mechanism to Aggregated Rate Law Mechanism. Rate vs substrate plot (a), and simulation from initial conditions (b)

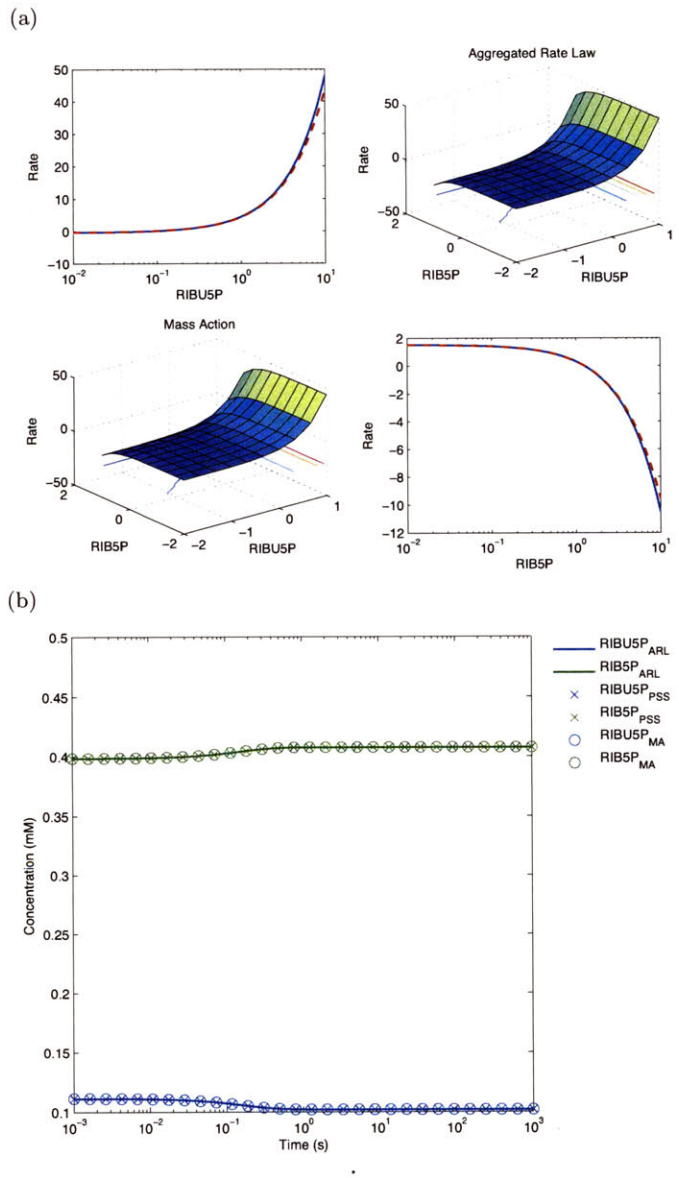


Figure B-19: Fit of R5P1 Mass Action Mechanism to Aggregated Rate Law Mechanism. Rate vs substrate plot (a), and simulation from initial conditions (b)

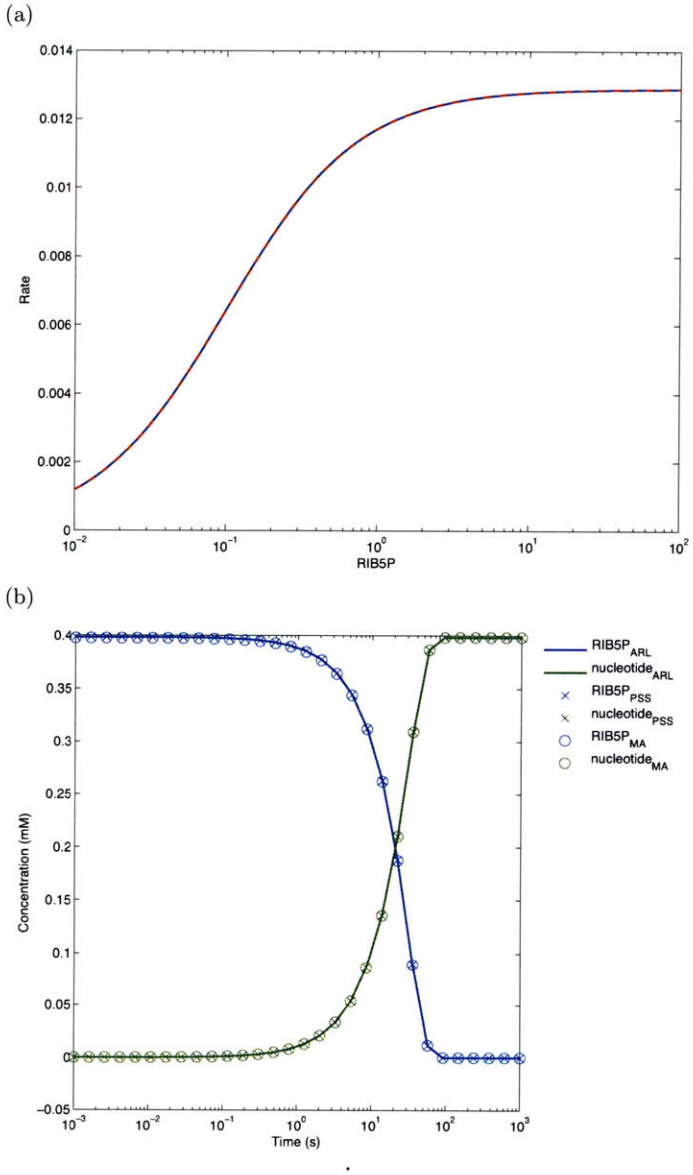


Figure B-20: Fit of RPPK Mass Action Mechanism to Aggregated Rate Law Mechanism. Rate vs substrate plot (a), and simulation from initial conditions (b)

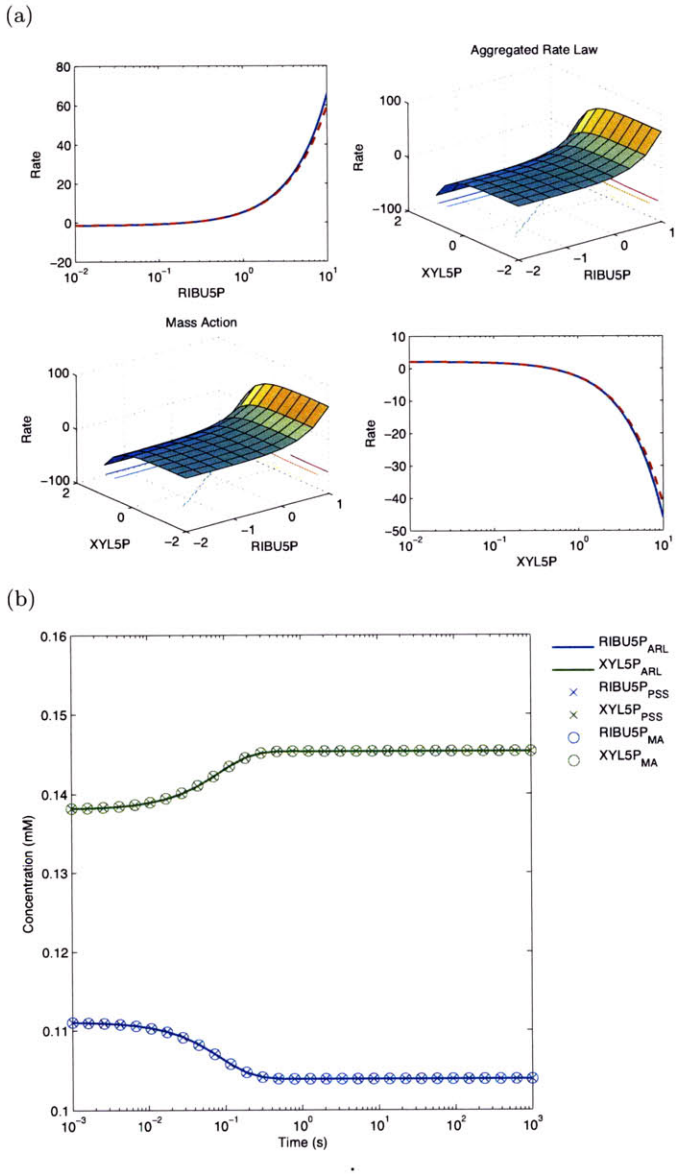


Figure B-21: Fit of Ru5P Mass Action Mechanism to Aggregated Rate Law Mechanism. Rate vs substrate plot (a), and simulation from initial conditions (b)

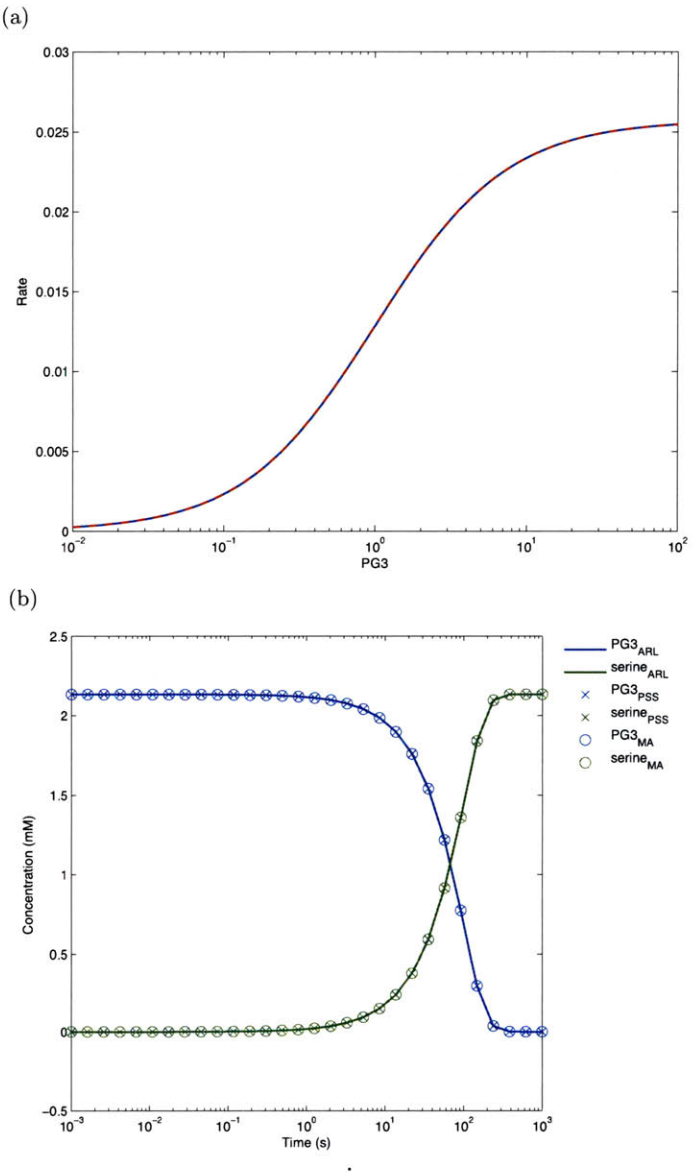


Figure B-22: Fit of SerSynth Mass Action Mechanism to Aggregated Rate Law Mechanism. Rate vs substrate plot (a), and simulation from initial conditions (b)

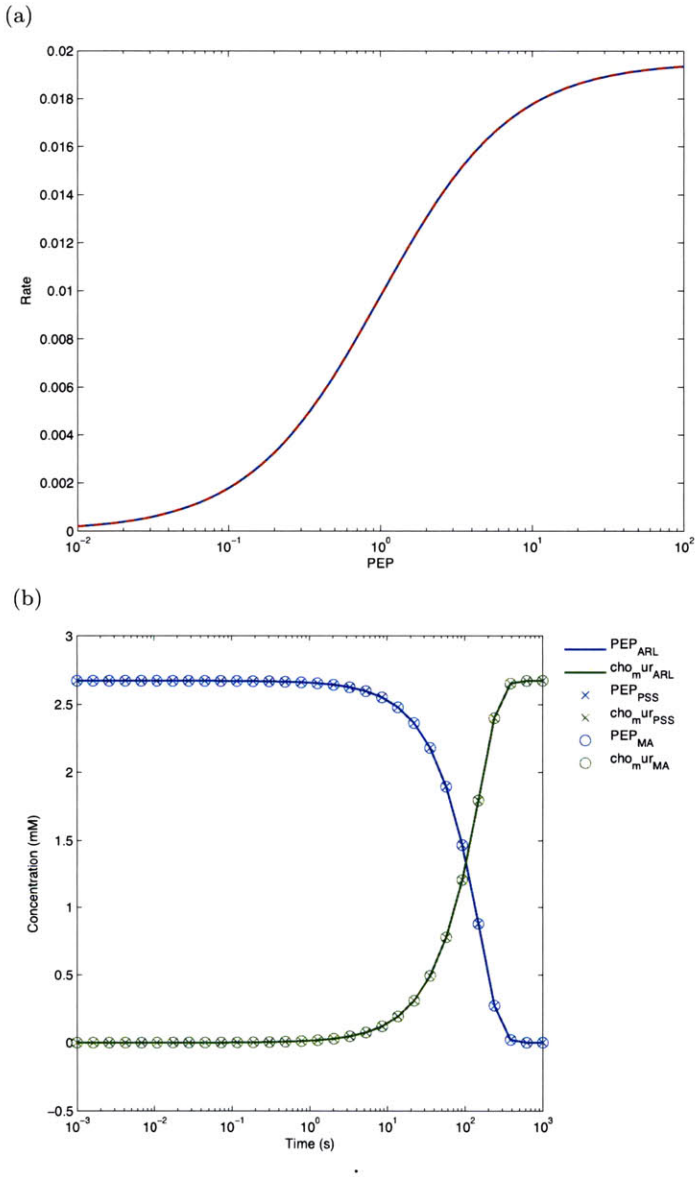


Figure B-23: Fit of Synth1 Mass Action Mechanism to Aggregated Rate Law Mechanism. Rate vs substrate plot (a), and simulation from initial conditions (b)

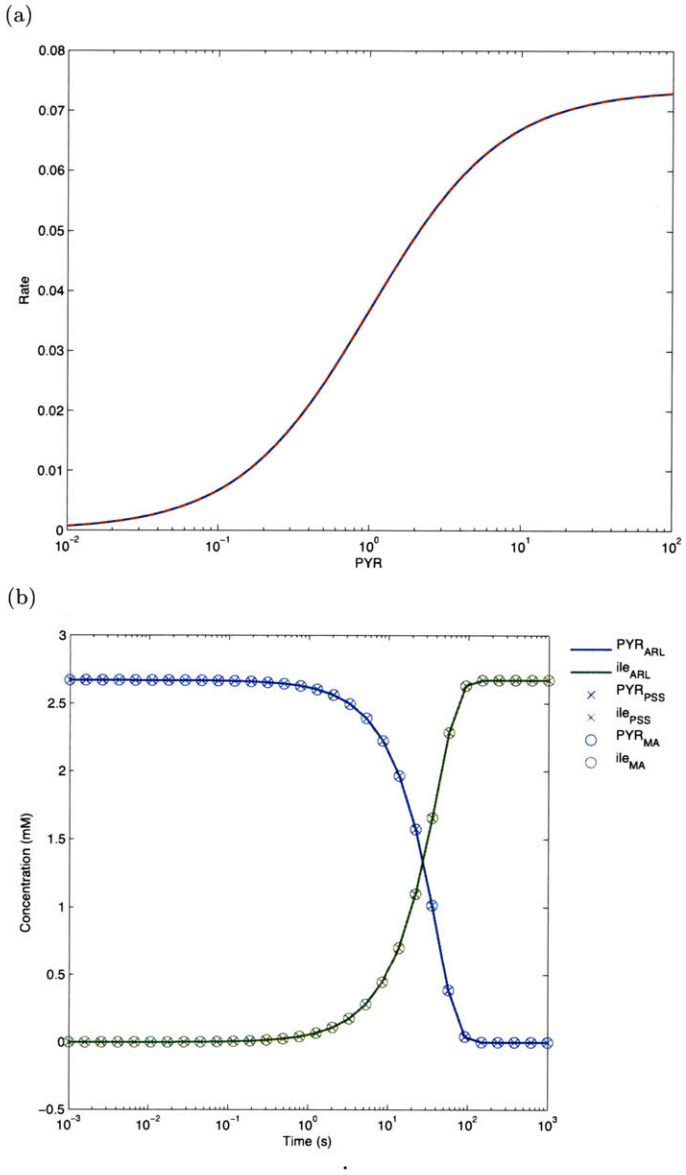
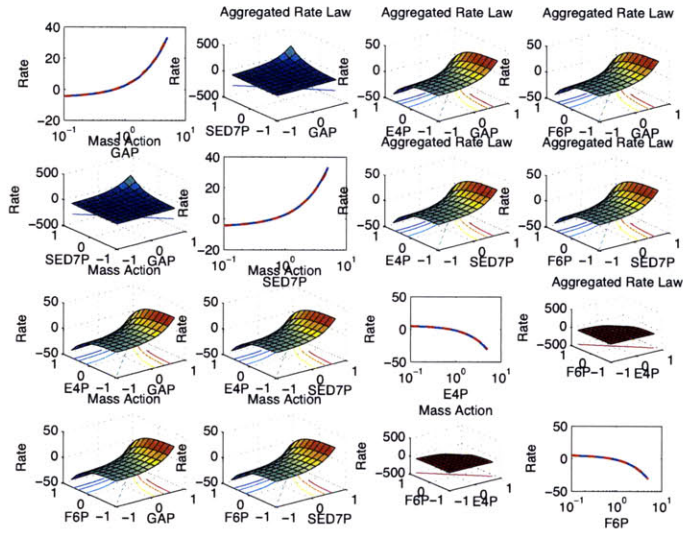


Figure B-24: Fit of Synth2 Mass Action Mechanism to Aggregated Rate Law Mechanism. Rate vs substrate plot (a), and simulation from initial conditions (b)

(a)



(b)

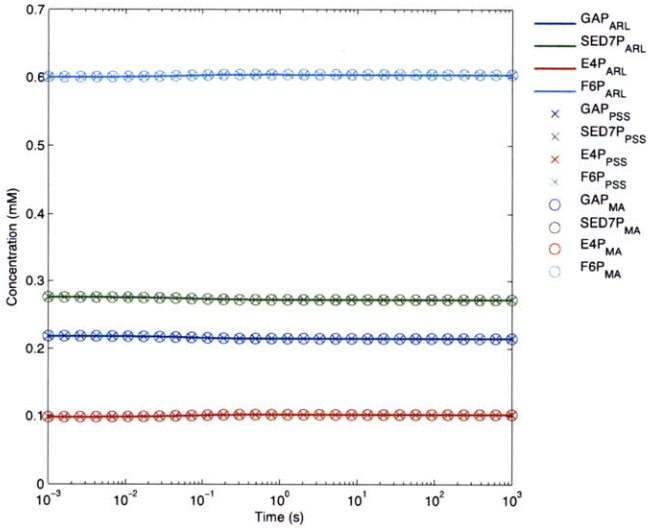


Figure B-25: Fit of TA Mass Action Mechanism to Aggregated Rate Law Mechanism. Rate vs substrate plot (a), and simulation from initial conditions (b)

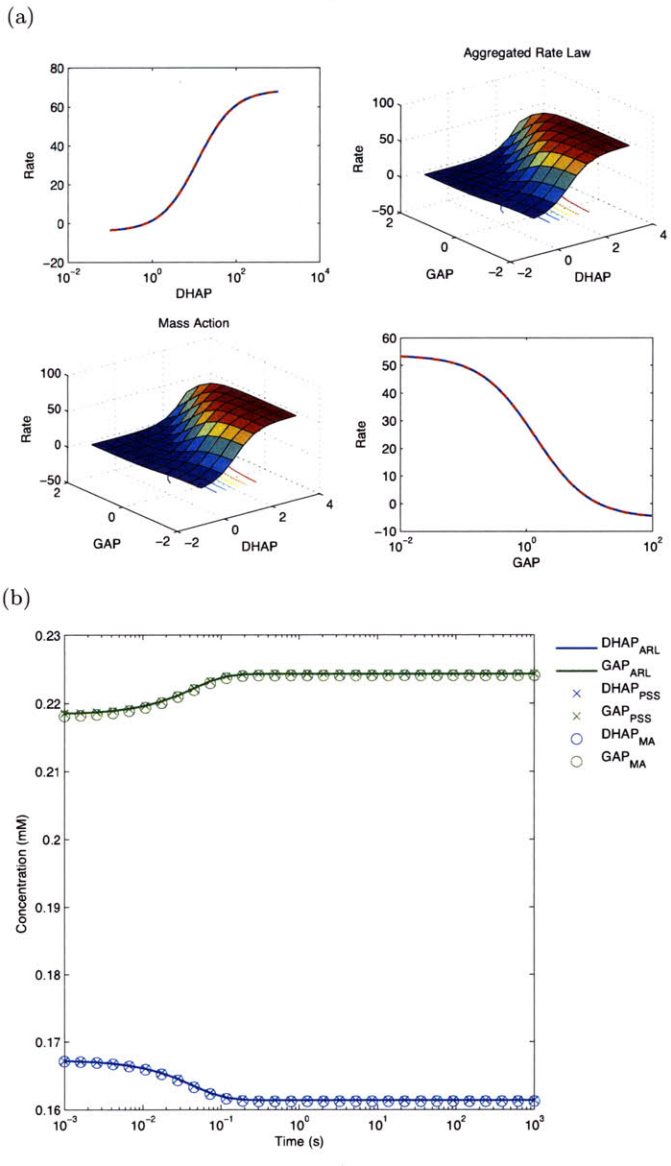
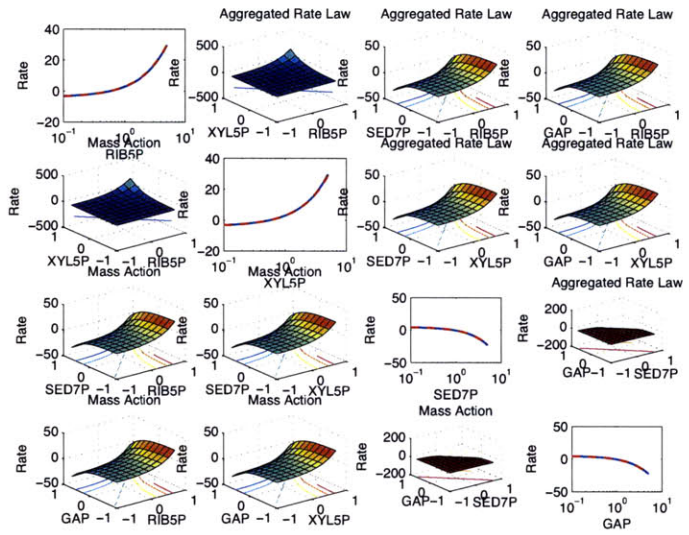


Figure B-26: Fit of TIS Mass Action Mechanism to Aggregated Rate Law Mechanism. Rate vs substrate plot (a), and simulation from initial conditions (b)

(a)



(b)

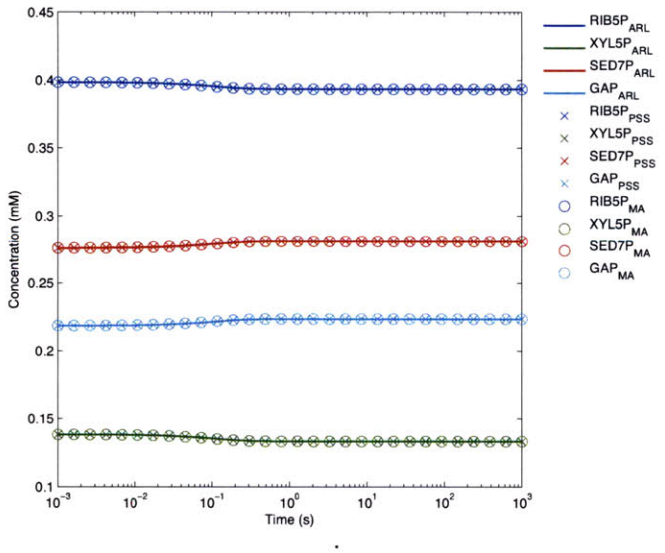
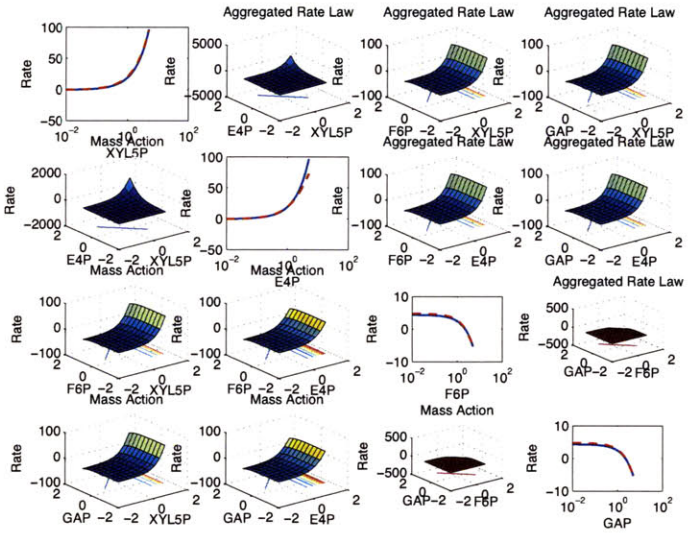


Figure B-27: Fit of TKa Mass Action Mechanism to Aggregated Rate Law Mechanism. Rate vs substrate plot (a), and simulation from initial conditions (b)

(a)



(b)

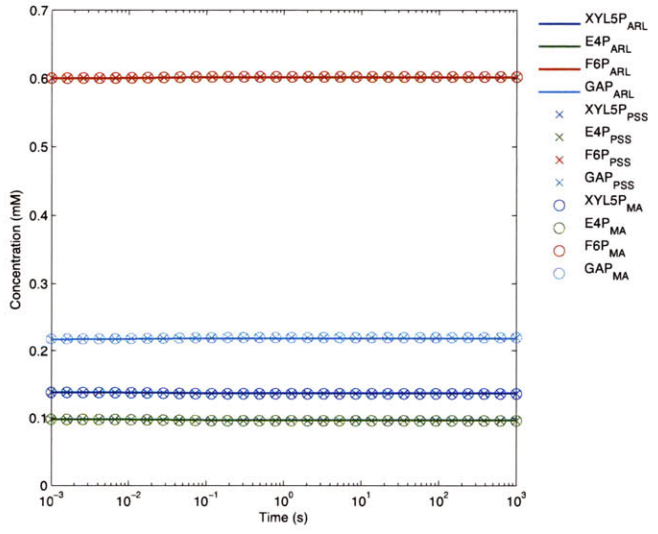


Figure B-28: Fit of TKb Mass Action Mechanism to Aggregated Rate Law Mechanism. Rate vs substrate plot (a), and simulation from initial conditions (b)

Bibliography

[1] B.S. Adiwijaya, P.I. Barton, and B. Tidor. Biological network design strategies: Discovery through dynamic optimization. *Molecular Biosystems*, 2(12):650–659, 2006.

[2] J.P. Akowski and R. Bauerle. Steady-State Kinetics and Inhibitor Binding of 3-Deoxy-d-arabino-heptulosonate-7-phosphate Synthase (Tryptophan sensitive) from Escherichia coli. *Biochemistry*, 36(50):15817–15822, 1997.

[3] I. Amit, A. Citri, T. Shay, Y. Lu, M. Katz, F. Zhang, G. Tarcic, D. Siwak, J. Lahad, J. Jacob-Hirsch, et al. A module of negative feedback regulators defines growth factor signaling. *Nature Genetics*, 39(4):503–512, 2007.

[4] M. Andrec, B.N. Kholodenko, R.M. Levy, and E. Sontag. Inference of signaling and gene regulatory networks by steady-state perturbation experiments: structure and accuracy. *Journal of Theoretical Biology*, 232(3):427–441, 2005.

[5] J.F. Apgar, J.E. Toettcher, D. Endy, F.M. White, and B. Tidor. Stimulus design for model selection and validation in cell signaling. *PLoS Computational Biology*, 4:e30, 2008.

[6] U.M. Ascher and L.R. Petzold. *Computer methods for ordinary differential equations and differential-algebraic equations.* Society for Industrial Mathematics, 1998.

[7] S.P. Asprey and S. Macchietto. Statistical tools for optimal dynamic model building. *Computers and Chemical Engineering*, 24(2-7):1261–1267, 2000.

[8] S.P. Asprey and S. Macchietto. Designing robust optimal dynamic experiments. *Journal of Process Control*, 12(4):545–556, 2002.

[9] C. Audet and J.E. Dennis Jr. Analysis of generalized pattern searches. *SIAM Journal on Optimization*, 13(3):889903, 2003.

[10] Y. Bard. *Nonlinear parameter estimation.* Academic Press New York, 1974.

[11] N. Barkai and S. Leibler. Robustness in simple biochemical networks. *Nature*, 387(6636):913–917, 1997.

[12] H. Barreteau, A. Kovac, A. Boniface, M. Sova, S. Gobec, and D. Blanot. Cytoplasmic steps of peptidoglycan biosynthesis. *FEMS Microbiology Reviews*, 32(2):168–207, 2008.

[13] J. Barthelmes, C. Ebeling, A. Chang, I. Schomburg, and D. Schomburg. BRENDA, AMENDA and FRENDA: the enzyme information system in 2007. *Nucleic Acids Research*, 35(Database issue):D511, 2007.

[14] D.P. Bertsekas. *Nonlinear Programming*. Athena Scientific, second edition edition, 1999.

[15] D.P. Bertsekas and J.N. Tsitsiklis. *Introduction to probability*. Athena Scientific, 2nd edition, 2008.

[16] C.A. Bever. *Selecting high-confidence predictions from ordinary differential equation models of biological networks*. PhD thesis, Massachusetts Institute of Technology, 2008.

[17] C. Birchmeier. Erbb receptors and the development of the nervous system. *Experimental Cell Research*, 315(4):611–8, Feb 15 2009.

[18] M.R. Birtwistle, M. Hatakeyama, N. Yumoto, B.A. Ogunnaike, J.B. Hoek, and B.N. Kholodenko. Ligand-dependent responses of the ErbB signaling network: experimental and modeling analyses. *Molecular Systems Biology*, 3(1), 2007.

[19] B. Blagoev, I. Kratchmarova, S.E. Ong, M. Nielsen, L.J. Foster, and M. Mann. A proteomics strategy to elucidate functional protein-protein interactions applied to egf signaling. *Nature Biotechnology*, 21(3):315–8, Mar 2003.

[20] N. Bluthgen, S. Legewie, S.M. Kielbasa, A. Schramme, O. Tchernitsa, J. Keil, A. Solf, M. Vingron, R. Schafer, H. Herzel, and C. Sers. A systems biological approach suggests that transcriptional feedback regulation by dual-specificity phosphatase 6 shapes extracellular signal-related kinase activity in RAS-transformed fibroblasts. *FEBS Journal*, 276(4):1024–1035, 2009.

[21] N. Borisov, E. Aksamitiene, A. Kiyatkin, S. Legewie, J. Berkhout, T. Maiwald, N.P. Kaimachnikov, J. Timmer, J.B. Hoek, and B.N. Kholodenko. Systems-level interactions between insulin–EGF networks amplify mitogenic signaling. *Molecular Systems Biology*, 5(1), 2009.

[22] P. Bork, L.J. Jensen, C. von Mering, A.K. Ramani, I. Lee, and E.M. Marcotte. Protein interaction networks from yeast to human. *Current Opinion in Structural Biology*, 14(3):292–299, 2004.

[23] G.E.P. Box and W.J. Hill. Discrimination among mechanistic models. *Technometrics*, 9(1):57–71, 1967.

[24] H. Bozdogan. Akaike's information criterion and recent developments in information complexity. *Journal of Mathematical Psychology*, 44(1):62–91, Mar 2000.

[25] R. Breitling and D. Hoeller. Current challenges in quantitative modeling of epidermal growth factor signaling. *FEBS Letters*, 579(28):6289–94, Nov 21 2005.

[26] G.E. Briggs and J.B.S. Haldane. A note on the kinetics of enzyme action. *The Biochemical Journal*, 19(2):338–339, 1925.

[27] F.A. Brightman and D.A. Fell. Differential feedback regulation of the MAPK cascade underlies the quantitative differences in EGF and NGF signalling in PC12 cells. *FEBS Letters*, 482(3):169–174, 2000.

[28] M. Brik Ternbach, C. Bollman, C. Wandrey, and R. Takors. Application of model discriminating experimental design for modeling and development of a fermentative fed-batch L-valine production process. *Biotechnology and Bioengineering*, 91(3), 2005.

[29] M. Brik Ternbach, C. Bollman, C. Wandrey, and R. Takors. Application of model discriminating experimental design for modeling and development of a fermentative fed-batch l-valine production process. *Biotechnology and Bioengineering*, 91(3):356–68, Aug 5 2005.

[30] S. Britsch. The neuregulin-i/erbb signaling system in development and disease. *Advances in Anatomy, Embryology, and Cell Biology*, 190:1–65, 2007.

[31] K. S. Brown, C. C. Hill, G. A. Calero, C. R. Myers, K. H. Lee, J. P. Sethna, and R. A. Cerione. The statistical mechanics of complex signaling networks: nerve growth factor signaling. *Physical Biology*, 1(3-4):184–95, Dec 2004.

[32] E. M. Bublil and Y. Yarden. The egf receptor family: spearheading a merger of signaling and therapeutics. *Current Opinion in Cell Biology*, 19(2):124–34, Apr 2007.

[33] S. Burden and Y. Yarden. Neuregulins and their receptors: A versatile signaling module in organogenesis and oncogenesis. *Neuron*, 18(6):847–55, Jun 1997.

[34] E.C. Butcher, E.L. Berg, and E.J. Kunkel. Systems biology in drug discovery. *Nature Biotechnology*, 22(10):1253–1259, 2004.

[35] G. Byrne and A. Hindmarsh. Stiff ODE solvers- A review of current and coming attractions. *Journal of Computational Physics*, 70(1):1–62, 1987.

[36] Y. Cao, S. T. Li, L. Petzold, and R. Serban. Adjoint sensitivity analysis or differential-algebraic equations: The adjoint dae system and its numerical solution. *Siam Journal on Scientific Computing*, 24(3):1076–1089, Jan 23 2003.

[37] K.D. Carey, A.J. Garton, M.S. Romero, J. Kahler, S. Thomson, S. Ross, F. Park, J.D. Haley, N. Gibson, and M.X. Sliwkowski. Kinetic analysis of epidermal growth factor receptor somatic mutant proteins shows increased sensitivity to the epidermal growth factor receptor tyrosine kinase inhibitor, erlotinib. *Cancer research*, 66(16):8163, 2006.

[38] F. P. Casey, D. Baird, Q. Feng, R. N. Gutenkunst, J. J. Waterfall, C. R. Myers, K. S. Brown, R. A. Cerione, and J. P. Sethna. Optimal experimental design in an epidermal growth factor receptor signalling and down-regulation model. *Systems Biology, IET*, 1(3):190–202, 2007.

[39] A. Chang, M. Scheer, A. Grote, I. Schomburg, and D. Schomburg. BRENDA, AMENDA and FRENDA the enzyme information system: new content and tools in 2009. *Nucleic Acids Research*, 2008.

[40] A. Chang, M. Scheer, A. Grote, I. Schomburg, and D. Schomburg. Brenda, amenda and frenda the enzyme information system: new content and tools in 2009. *Nucleic Acids Research*, 37(suppl. 1):D588–592, 2009.

[41] C. Chassagnole, N. Noisommit-Rizzi, J.W. Schmid, K. Mauch, and M. Reuss. Dynamic modeling of the central carbon metabolism of Escherichia coli. *Biotechnology and Bioengineering*, 79(1), 2002.

[42] B.H. Chen and S.P. Asprey. On the design of optimally informative dynamic experiments for model discrimination in multiresponse nonlinear situations. *Industrial and Engineering Chemistry Research*, 42(7):1379–1390, 2003.

[43] W.W. Chen, B. Schoeberl, P.J. Jasper, M. Niepel, U.B. Nielsen, D.A. Lauffenburger, and P.K. Sorger. Input–output behavior of ErbB signaling pathways as revealed by a mass action model trained against dynamic data. *Molecular Systems Biology*, 5(1), 2009.

[44] K. Cho, S. Shin, W. Kolch, and O.Wolkenhauer. Experimental design in systems biology, based on parameter sensitivity analysis using a monte carlo method: A case study for the tnfalpha-mediated nf-kappa b signal transduction pathway. *Simulation*, 79(12):726–739, 2003.

[45] A. Citri and Y. Yarden. EGF–ERBB signalling: towards the systems level. *Nature Reviews Molecular Cell Biology*, 7(7):505–516, 2006.

[46] W.W. Cleland. The kinetics of enzyme-catalyzed reactions with two or more substrates or products. I. Nomenclature and rate equations. *Biochimica et Biophysica Acta*, 67:104, 1963.

[47] C. Cobelli and J.J. DiStefano. Parameter and structural identifiability concepts and ambiguities: a critical review and analysis. *American Journal of Physiology- Regulatory, Integrative and Comparative Physiology*, 239(1):7–24, 1980.

[48] M. Cooney and K. McDonald. Optimal dynamic experiments for bioreactor model discrimination. *Applied Microbiology and Biotechnology*, 43(5):826–837, 1995.

[49] A. Cornish-Bowden. An Automatic Method for Deriving Steady-State Rate Equations. *The Biochemical Journal*, 165:55–59, 1977.

[50] H. Dahari, J.E. Layden-Almer, E. Kallwitz, R.M. Ribeiro, S.J. Cotler, T.J. Layden, and A.S. Perelson. A mathematical model of hepatitis c virus dynamics in patients with high baseline viral loads or advanced liver disease. *Gastroenterology*, 136(4):1402–9, Apr 2009.

[51] M.D. Deng, A.D. Grund, S.L. Wassink, S.S. Peng, K.L. Nielsen, B.D. Huckins, B.L. Walsh, and R.P. Burlingame. Directed evolution and characterization of Escherichia coli glucosamine synthase. *Biochimie*, 88(5):419–429, 2006.

[52] B.M.R. Donckels, D.J.W. De Pauw, P.A. Vanrolleghem, and B. De Baets. A kernel-based method to determine optimal sampling times for the simultaneous estimation of the parameters of rival mathematical models. *Journal of Computational Chemistry*, 2009.

[53] A. Eastman and R.P. Perez. New targets and challenges in the molecular therapeutics of cancer. *British Journal of Clinical Pharmacology*, 62(1):5–14, Jul 2006.

[54] L. Endler, N. Rodriguez, N. Juty, V. Chelliah, C. Laibe, C. Li, and N. Le Novère. Designing and encoding models for synthetic biology. *Journal of the Royal Society Interface*, 2009.

[55] J.M. English and M.H. Cobb. Pharmacological inhibitors of mapk pathways. *Trends in Pharmacological Sciences*, 23(1):40 – 45, 2002.

[56] D. Faller, U. Klingmuller, and J. Timmer. Simulation methods for optimal experimental design in systems biology. *Simulation*, 79(12):717–725, 2003.

[57] E.M. Fallon and Douglas A. Lauffenburger. Computational model for effects of ligand/receptor binding properties on interleukin-2 trafficking dynamics and t cell proliferation response. *Biotechnology Progress*, 16(5):905–916, 2000.

[58] I. Famili and B.O. Palsson. The convex basis of the left null space of the stoichiometric matrix leads to the definition of metabolically meaningful pools. *Biophysical Journal*, 85(1):16–26, Jul 2003.

[59] J.E. Ferrell and R.R. Bhatt. Mechanistic studies of the dual phosphorylation of mitogen-activated protein kinase. *The Journal of Biological Chemistry*, 272(30):19008–19016, 1997.

[60] A. Fersht. *Structure and Mechanism in Protein Science.* W. H. Freeman and Company, 1999.

[61] S. Fields and O. Song. A novel genetic system to detect protein protein interactions. *Nature*, 340(6230):245–246, 1989.

[62] A. Finney and M. Hucka. Systems biology markup language: Level 2 and beyond. *Biochemical Society Transactions*, 31(Pt 6):1472–3, Dec 2003.

[63] J.B. Fitzgerald, B. Schoeberl, U.B. Nielsen, and P.K. Sorger. Systems biology and combination therapy in the quest for clinical efficacy. *Nature Chemical Biology*, 2(9):458–466, 2006.

[64] P. Flaherty, M. L. Radhakrishnan, T. Dinh, R.A. Rebres, T.I. Roach, M.I. Jordan, and A.P. Arkin. A dual receptor crosstalk model of g-protein-coupled signal transduction. *PLoS Computational Biology*, 4(9):e1000185, 2008.

[65] N. Friedman. Inferring cellular networks using probabilistic graphical models. *Science*, 303(5659):799–805, 2004.

[66] A. Fujioka, K. Terai, R.E. Itoh, K. Aoki, T. Nakamura, S. Kuroda, E. Nishida, and M. Matsuda. Dynamics of the Ras/ERK MAPK Cascade as Monitored by Fluorescent Probes. *The Journal of Biological Chemistry*, 281(13):8917–8926, 2006.

[67] K.G. Gadkar, R. Gunawan, and F.J. Doyle. Iterative approach to model identification of biological networks. *BMC Bioinformatics*, 6(1):155, 2005.

[68] D. Garfinkel, C.B. Marbach, and N.Z. Shapiro. Stiff differential equations. *Annual Review of Biophysics and Bioengineering*, 6(1):525–542, 1977.

[69] N. Geva-Zatorsky, N. Rosenfeld, S. Itzkovitz, R. Milo, A. Sigal, E. Dekel, T. Yarnitzky, Y. Liron, P. Polak, G. Lahav, and U. Alon. Oscillations and variability in the p53 system. *Molecular Systems Biology*, 2(1), 2006.

[70] P.E. Gill, W. Murray, and M.H. Wright. *Practical Optimization.* Academic Press, 1981.

[71] D.T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22(4):403–434, 1976.

[72] D.T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, 1977.

[73] A. Goldbeter and D. E. Koshland Jr. An amplified sensitivity arising from covalent modification in biological systems. *Proceedings of the National Academy of Sciences of the United States of America*, 78(11):6840–4, Nov 1981.

[74] A. Goldbeter and L.A. Segel. Unified mechanism for relay and oscillation of cyclic AMP in Dictyostelium discoideum. *Proceedings of the National Academy of Sciences of the United States of America*, 74(4):1543–1547, 1977.

[75] A. Graham. *Kronecker products and matrix calculus with applications.* Halsted Press, 1981.

[76] A. Gschwind, O.M. Fischer, and A. Ullrich. The discovery of receptor tyrosine kinases: targets for cancer therapy. *Nature Reviews Cancer*, 4(5):361–370, 2004.

[77] W. Gu, X. Zhu, N. Futai, B.S. Cho, and S. Takayama. Computerized microfluidic cell culture using elastomeric channels and Braille displays. *Proceedings of the National Academy of Sciences of the United States of America*, 101(45):15861–15866, 2004.

[78] R.N. Gutenkunst, J.J. Waterfall, F.P. Casey, K.S. Brown, C.R. Myers, and J.P. Sethna. Universally sloppy parameter sensitivities in systems biology models. *PLoS Computational Biology*, 3(10), 2007.

[79] L. Hakes, J.W. Pinney, D.L. Robertson, and S.C. Lovell. Protein-protein interaction networks and biology–what's the connection? *Nature Biotechnology*, 26(1):69–72, Jan 2008.

[80] J.M. Haugh. Mathematical model of human growth hormone (hgh)-stimulated cell proliferation explains the efficacy of hgh variants as receptor agonists or antagonists. *Biotechnology Progress*, 20(5):1337–44, Sep-Oct 2004.

[81] M.D. Haunschild, B. Freisleben, R. Takors, and W. Wiechert. Investigating the dynamic behavior of biochemical networks using model families. *Bioinformatics*, 21(8):1617–25, Apr 15 2005.

[82] E.B. Haura, J. Turkson, and R. Jove. Mechanisms of disease: Insights into the emerging role of signal transducers and activators of transcription in cancer. *Nature Clinical Practice Oncology*, 2(6):315–324, 2005.

[83] G.J. Heisermann, H.S. Wiley, B.J. Walsh, H.A. Ingraham, C.J. Fiol, and G.N. Gill. Mutational removal of the Thr669 and Ser671 phosphorylation sites alters substrate specificity and ligand-induced internalization of the epidermal growth factor receptor. *Journal of Biological Chemistry*, 265(22):12820–12827, 1990.

[84] S. Hengl, C. Kreutz, J. Timmer, and T. Maiwald. Data-based identifiability analysis of non-linear dynamical models. *Bioinformatics*, 23(19):2612, 2007.

[85] E. Hofmann and G. Kopperschläger. Phosphofructokinase from yeast. *Methods in Enzymology*, 90:49, 1982.

[86] J.J. Hornberg, B. Binder, F.J. Bruggeman, B. Schoeberl, R. Heinrich, and H.V. Westerhoff. Control of mapk signalling: From complexity to what really matters. *Oncogene*, 24(36):5533–42, Aug 25 2005.

[87] C.Y. Huang and J.E. Ferrell Jr. Ultrasensitivity in the mitogen-activated protein kinase cascade. *Proceedings of the National Academy of Sciences of the United States of America*, 93(19):10078–83, Sep 17 1996.

[88] M. Hucka, A. Finney, H. M. Sauro, H. Bolouri, J. C. Doyle, H. Kitano, A. P. Arkin, B. J. Bornstein, D. Bray, A. Cornish-Bowden, A. A. Cuellar, S. Dronov, E. D. Gilles, M. Ginkel, V. Gor, II Goryanin , W. J. Hedley, T. C. Hodgman, J. H. Hofmeyr, P. J. Hunter, N. S. Juty, J. L. Kasberger, A. Kremling, U. Kummer, N. Le Novere, L. M. Loew, D. Lucio, P. Mendes, E. Minch, E. D. Mjolsness, Y. Nakayama, M. R. Nelson, P. F. Nielsen, T. Sakurada, J. C. Schaff, B. E. Shapiro, T. S. Shimizu, H. D. Spence, J. Stelling, K. Takahashi, M. Tomita, J. Wagner, and J. Wang. The systems biology markup language (sbml): a medium for representation and exchange of biochemical network models. *Bioinformatics*, 19(4):524–31, Mar 1 2003.

[89] K. Jaqaman and G. Danuser. Linking data to models: Data regression. *Nature reviews Molecular Cell Biology*, 7(11):813–9, Nov 2006.

[90] R.B. Jones, A. Gordus, J.A. Krall, and G. MacBeath. A quantitative protein interaction network for the ErbB receptors using protein microarrays. *Nature*, 439(7073):168–174, 2005.

[91] E.J. Joslin, L.K. Opresko, A. Wells, H.S. Wiley, and D.A. Lauffenburger. EGF-receptor-mediated mammary epithelial cell migration is driven by sustained ERK signaling from autocrine stimulation. *Journal of Cell Science*, 120(20):3688, 2007.

[92] E.H. Kaplan, D.L. Craft, and L.M. Wein. Emergency response to a smallpox attack: the case for mass vaccination. *Proceedings of the National Academy of Sciences of the United States of America*, 99(16):10935–10940, 2002.

[93] I.M. Keseler, J. Collado-Vides, S. Gama-Castro, J. Ingraham, S. Paley, I.T. Paulsen, M. Peralta-Gil, and P.D. Karp. EcoCyc: a comprehensive database resource for Escherichia coli. *Nucleic Acids Research*, 33(Database Issue):D334, 2005.

[94] Z. Khan, V. Pillay, Y.E. Choonara, and L.C. du Toit. Drug delivery technologies for chronotherapeutic applications. *Pharmaceutical Development and Technology*, pages 1–12, 2009.

[95] B.N. Kholodenko, O.V. Demin, G. Moehren, and J.B. Hoek. Quantification of short term signaling by the epidermal growth factor receptor. *The Journal of Biological Chemistry*, 274(42):30169–30181, 1999.

[96] E.L. King and C. Altman. A schematic method of deriving the rate laws for enzyme-catalyzed reactions. *The Journal of Physical Chemistry*, 60(10):1375–1378, 1956.

[97] H. Kitano. Systems biology : A brief overview. *Science*, 295(5560):1662–4, Mar 1 2002.

[98] H. Kitano. A robustness-based approach to systems-oriented drug design. *Nature Reviews Drug Discovery*, 6(3):202–210, 2007.

[99] Albe K.R., Butler M.H., and Wright B.E. Cellular concentrations of enzymes and their substrates. *Journal of Theoretical Biology*, 22143(2):163–95, 1990.

[100] A. Kremling, S. Fischer, K. Gadkar, F.J. Doyle, T. Sauter, E. Bullinger, F. Allgower, and E.D. Gilles. A benchmark for methods in reverse engineering and model discrimination: Problem formulation and solutions. *Genome Research*, 14(9):1773–1785, 2004.

[101] A. Kremling and J. Saez-Rodriguez. Systems biology: An engineering perspective. *Journal of Biotechnology*, 129(2):329–351, 2007.

[102] N. Kumar, B.S. Hendriks, K.A. Janes, D. de Graaf, and D.A. Lauffenburger. Applying computational modeling to drug discovery and development. *Drug discovery today*, 11(17-18):806–811, 2006.

[103] Z. Kutalik, Kwang-Hyun Cho, and O. Wolkenhauer. Optimal sampling time selection for parameter estimation in dynamic pathway modeling. *Biosystems*, 75(1-3):43–55, 2004.

[104] P. Kuzmic. The king-altman method. `http://www.biokin.com/king-altman/index.html`.

[105] K.J. Laidler. Chemical kinetics and the origins of physical chemistry. *Archive for History of Exact Sciences*, 32(1):43–75, 1985.

[106] N. Le Novere, B. Bornstein, A. Broicher, M. Courtot, M. Donizelli, H. Dharuri, L. Li, H. Sauro, M. Schilstra, B. Shapiro, J. L. Snoep, and M. Hucka. Biomodels database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Research*, 34(Database issue):D689–91, Jan 1 2006.

[107] R.M. Lewis and V. Torczon. Pattern search algorithms for bound constrained minimization. *SIAM Journal on Optimization*, 9(4):10821099, 1999.

[108] R.M. Lewis and V. Torczon. Pattern search methods for linearly constrained minimization. *SIAM Journal on Optimization*, 10(3):917941, 2000.

[109] R.M. Lewis and V. Torczon. A globally convergent augmented lagrangian pattern search algorithm for optimization with general constraints and simple bounds. *SIAM Journal on Optimization*, 12(4):10751089, 2002.

[110] J.C. Liao, S.Y. Hou, and Y.P. Chao. Pathway analysis, engineering, and physiological considerations for redirecting central metabolism. *Biotechnology and Bioengineering*, 52(1), 1996.

[111] C. A. Lipschultz, Y. Li, and S. Smith-Gill. Experimental design for analysis of complex kinetics using surface plasmon resonance. *Methods*, 20(3):310–8, Mar 2000.

[112] C.M. Lloyd, M.D. Halstead, and P.F. Nielsen. Cellml: its future, present and past. *Progress in Biophysics and Molecular Biology*, 85(2-3):433–50, Jun-Jul 2004.

[113] H. Lodish, A. Berk, L.S. Zipursky, P. Matsudaira, D. Baltimore, and J. Darnell. *Molecular Cell Biology*. W.H. Freeman and Company, 4th edition, 2000.

[114] S. Marsili-Libelli, S. Guerrizio, and N. Checchi. Confidence regions of estimated parameters for ecological systems. *Ecological Modelling*, 165(2-3):127 – 146, 2003.

[115] J. Martins, P. Sturdza, and J.J. Alonso. The complex-step derivative approximation. *ACM Transactions on Mathematical Software (TOMS)*, 29(3):245–262, 2003.

[116] J.M. Mason, D.J. Morrison, M. Albert Basson, and J.D. Licht. Sprouty proteins: multifaceted negative-feedback regulators of receptor tyrosine kinase signaling. *Trends in Cell Biology*, 16(1):45–54, 2006.

[117] M.A. Matthay, J.P. Thiery, F. Lafont, F. Stampfer, and B. Boyer. Transient effect of epidermal growth factor on the motility of an immortalized mammary epithelial cell line. *Journal of Cell Science*, 106(3):869–878, 1993.

[118] Franklin C. McLean. Application of the law of chemical equilibrium (law of mass action) to biological problems. *Physiological Reviews*, 18(4):495–523, 1938.

[119] A.D. McNaught and A. Wilkinson. *IUPAC compendium of chemical terminology.* IUPAC, 2nd edition, 1997.

[120] D.A. McQuarrie. Stochastic approach to chemical kinetics. *Journal of Applied Probability*, pages 413–478, 1967.

[121] Stroppolo M.E., Falconi M., Caccuri A.M., and Desideri A. Superefficient enzymes. *Cellular and Molecular Life Sciences CMLS*, 58(10):1451–60, 2001.

[122] J. Mendelsohn and J. Baselga. The egf receptor family as targets for cancer therapy. *Oncogene*, 19(56):6550–65, Dec 27 2000.

[123] L. Michaelis and M.L. Menten. Kinetics of invertase action. *Biochemistry Z*, 49:333–369, 1913.

[124] N. Moghal and P.W. Sternberg. Multiple positive and negative regulators of signaling by the EGF-receptor. *Current Opinion In Cell Biology*, 11(2):190–196, 1999.

[125] C.G. Moles, P. Mendes, and J.R. Banga. Parameter estimation in biochemical pathways: a comparison of global optimization methods. *Genome Research*, 13(11):2467–74, Nov 2003.

[126] R. Mulloy, A. Ferrand, Y. Kim, R. Sordella, D.W. Bell, D.A. Haber, K.S. Anderson, and J. Settleman. Epidermal growth factor receptor mutants from human lung cancers exhibit enhanced catalytic activity and increased sensitivity to gefitinib. *Cancer Research*, 67(5):2325, 2007.

[127] A. Munack. *Some improvements in the identification of bioprocesses.* IFAC Symposia Series. Pergamon Press, 1992.

[128] J.R.S. Newman and A.E. Keating. Comprehensive identification of human bZIP interactions with coiled-coil arrays. *Science*, 300(5628):2097–2101, 2003.

[129] D. Nickerson, C. Stevens, M. Halstead, P. Hunter, and P. Nielsen. Toward a curated cellml model repository. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, 1:4237–40, 2006.

[130] U.B. Nielsen and B. Schoeberl. Using computational modeling to drive the development of targeted therapeutics. *IDrugs the Investigational Drugs Journal*, 8(10):822, 2005.

[131] J. Nocedal and S.J. Wright. *Numerical Optimization.* Springer Verlag, second edition edition, 2006.

[132] K. Oda, Y. Matsuoka, A. Funahashi, and H. Kitano. A comprehensive pathway map of epidermal growth factor receptor signaling. *Molecular systems biology*, 1:2005 0010, 2005.

[133] J.V. Olsen, B. Blagoev, F. Gnad, B. Macek, C. Kumar, P. Mortensen, and M. Mann. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell*, 127(3):635–48, Nov 3 2006.

[134] M. Ono, A. Hirata, T. Kometani, M. Miyagawa, S. Ueda, H. Kinoshita, T. Fujii, and M. Kuwano. Sensitivity to gefitinib (Iressa, ZD1839) in non-small cell lung cancer cell lines correlates with dependence on the epidermal growth factor (EGF) receptor/extracellular signal-regulated kinase 1/2 and EGF receptor/Akt pathway for proliferation. *Molecular Cancer Therapeutics*, 3(4):465–472, 2004.

[135] A.V. Oppenheim, A.S. Willsky, and S.H. Nawab. *Signals and systems*. Prentice-Hall, Inc., 1996.

[136] D.B. Özyurt and P.I. Barton. Cheap second order directional derivatives of stiff ode embedded functionals. *Siam Journal on Scientific Computing*, 26(5):1725–1743, 2005.

[137] D.B. Özyurt and P.I. Barton. Large-scale dynamic optimization using the directional second-order adjoint method. *Industrial and Engineering Chemistry Research*, 44(6):1804–1811, 2005.

[138] J.G. Paez, P.A. Janne, J.C. Lee, S. Tracy, H. Greulich, S. Gabriel, P. Herman, F.J. Kaye, N. Lindeman, T.J. Boggon, K. Naoki, H. Sasaki, Y. Fujii, M.J. Eck, W.R. Sellers, B.E. Johnson, and M. Meyerson. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science*, 304(5676):1497–1500, 2004.

[139] A.S. Perelson, A.U. Neumann, M. Markowitz, J.M. Leonard, and D.D. Ho. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science*, 271(5255):1582, 1996.

[140] N. Perrimon and A.P. McMahon. Negative Feedback Mechanisms Minireview and Their Roles during Pattern Formation. *Cell*, 97:13–16, 1999.

[141] L.M. Prisant. Chronotherapeutics: A surge of ideas. *Clinical Cornerstone*, 6(4):7–15, 2004.

[142] H. Rabitz, M. Kramer, and D. Dacol. Sensitivity analysis in chemical kinetics. *Annual review of physical chemistry*, 34(1):419–461, 1983.

[143] G.T. Reeves, C.B. Muratov, T. Sch
"upbach, and S.Y. Shvartsman. Quantitative models of developmental pattern formation. *Developmental cell*, 11(3):289–300, 2006.

[144] M. Rodriguez-Fernandez, P. Mendes, and J. R. Banga. A hybrid approach for efficient and robust parameter estimation in biochemical pathways. *Biosystems*, 83(2-3):248–65, Feb-Mar 2006.

[145] D.M. Rothman, M.D. Shults, and B. Imperiali. Chemical approaches for investigating phosphorylation in signal transduction networks. *Trends in Cell Biology*, 15(9):502–510, 2005.

[146] C. Rubin, V. Litvak, H. Medvedovsky, Y. Zwang, S. Lev, and Y. Yarden. Sprouty fine-tunes EGF signaling through interlinked positive and negative feedback loops. *Current Biology*, 13(4):297–307, 2003.

[147] K. Sachs, D. Gifford, T. Jaakkola, P. Sorger, and D.A. Lauffenburger. Bayesian network approach to cell signaling pathway modeling. *Science's STKE*, 2002(148), 2002.

[148] A. Saltelli, M. Ratto, S. Tarantola, and F. Campolongo. Sensitivity analysis for chemical models. *Chemical Reviews*, 105(7):2811–2828, 2005.

[149] S. Sasagawa, Y. Ozaki, K. Fujita, and S. Kuroda. Prediction and validation of the distinct dynamics of transient and sustained ERK activation. *Nature Cell Biology*, 7(4):365–373, 2005.

[150] H.M. Sauro and B. Ingalls. Conservation analysis in biochemical networks: computational issues for software writers. *Biophysical Chemistry*, 109(1):1–15, 2004.

[151] C.L. Sawyers. Rational therapeutic intervention in cancer: kinases as drug targets. *Current Opinion in Genetics & Development*, 12(1):111–115, 2002.

[152] B. Schoeberl, C. Eichler-Jonsson, E.D. Gilles, and G. Muller. Computational modeling of the dynamics of the map kinase cascade activated by surface and internalized egf receptors. *Nature Biotechnology*, 20(4):370–375, 2002.

[153] G.A.F. Seber and C.J. Wild. *Nonlinear regression*. Wiley-IEEE, 2003.

[154] I.H. Segel. *Enzyme kinetics. New York*. NY John Wiley and Sons, 1975.

[155] L.A. Segel. On the validity of the steady state assumption of enzyme kinetics. *Bulletin of Mathematical Biology*, 50(6):579–593, 1988.

[156] L.F Shampine and M.W. Reichelt. The matlab ode suite. *Siam Journal on Scientific Computing*, 18:1–22, 1997.

[157] Y.D. Shaul and R. Seger. The mek/erk cascade: From signaling specificity to diverse functions. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1773(8):1213 – 1226, 2007. Mitogen-Activated Protein Kinases: New Insights on Regulation, Function and Role in Human Disease.

[158] L.K. Shawver, D. Slamon, and A. Ullrich. Smart drugs: tyrosine kinase inhibitors in cancer therapy. *Cancer Cell*, 1(2):117–123, 2002.

[159] S. Shimizu-Sato, E. Huq, J.M. Tepperman, and P.H. Quail. A light-switchable gene promoter system. *Nature Biotechnology*, 20(10):1041–1044, 2002.

[160] V.K. Singh and I. Ghosh. Kinetic modeling of tricarboxylic acid cycle and glyoxylate bypass in Mycobacterium tuberculosis, and its application to assessment of drug targets. *Theoretical Biology and Medical Modelling*, 3(1):27, 2006.

[161] E. Sontag, A. Kiyatkin, and B.N. Kholodenko. Inferring dynamic architecture of cellular networks using time series of gene expression, protein and metabolite data. *Bioinformatics*, 20(12):1877, 2004.

[162] A.B. Sprenkle, S.P. Davies, D. Carling, D.G. Hardie, and T.W. Sturgill. Identification of Raf-1 Ser621 kinase activity from NIH 3T3 cells as AMP-activated protein kinase. *FEBS Letters*, 403(3):254, 1997.

[163] M.R. Stampfer and J.C. Bartley. Induction of transformation and continuous cell lines from normal human mammary epithelial cells after exposure to benzo [a] pyrene. *Proceedings of the National Academy of Sciences of the United States of America*, 82(8):2394–2398, 1985.

[164] H. Steven Wiley, S.Y. Shvartsman, and D.A. Lauffenburger. Computational modeling of the EGF-receptor system: a paradigm for systems biology. *Trends in Cell Biology*, 13(1):43–50, 2003.

[165] G. Strang. *Introduction to Applied Mathematics*. Wellesley-Cambridge Press, 1986.

[166] A.I. Su, T. Wiltshire, S. Batalov, H. Lapp, K.A. Ching, D. Block, J. Zhang, R. Soden, M. Hayakawa, G. Kreiman, M.P. Cooke, J.R. Walker, and J.B. Hogenesch. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proceedings of the National Academy of Sciences of the United States of America*, 101(16):6062–7, Apr 20 2004.

[167] K. Takishima, B.A. Friedman, H. Fujiki, and M. Rich Rosner. Thapsigargin, a novel promoter, phosphorylates the epidermal growth factor receptor at threonine 669. *Biochemical and biophysical research communications(Print)*, 157(2):740–746, 1988.

[168] The Mathworks, Inc. MATLAB(tm) 2007a, 2007.

[169] The Mathworks, Inc. MATLAB(tm) 2008b, 2008.

[170] J.E. Toettcher, A. Loewer, G.J. Ostheimer, M.B. Yaffe, B. Tidor, and G. Lahav. Distinct mechanisms act in concert to mediate cell cycle arrest. *Proceedings of the National Academy of Sciences of the United States of America*, 106(3):785, 2009.

[171] V. Torczon. On the convergence of pattern search algorithms. *SIAM Journal on Optimization*, 7(1):125, 1997.

[172] A.R. Tzafriri. Michaelis-Menten kinetics at high enzyme concentrations. *Bulletin of Mathematical Biology*, 65(6):1111–1129, 2003.

[173] N.A.W. van Riel. Dynamic modelling and analysis of biochemical networks: mechanism-based models and model-based experiments. *Briefings in Bioinformatics*, 7(4):364, 2006.

[174] N.A.W. van Riel and E.D. Sontag. Parameter estimation in models combining signal transduction and metabolic pathways: the dependent input approach. *Systems Biology, IEE Proceedings*, 153(4):263–274, 2006.

[175] F.G. Vital-Lopez, A. Armaou, E.V. Nikolaev, and C.D. Maranas. A computational procedure for optimal engineering interventions using kinetic models of metabolism. *Biotechnology progress*, 22(6), 2006.

[176] C. von Mering, R. Krause, B. Snel, M. Cornell, S.G. Oliver, S. Fields, and P. Bork. Comparative assessment of large-scale data sets of protein–protein interactions. *Nature*, 417(6887):399–403, 2002.

[177] D. von Seggern. *CRC Standard Curves and Surfaces*. CRC Press, 1993.

[178] P. Waage and C.M. Guldberg. Studies concerning affinity. *Forhandlinger: Videnskabs-Selskabet i Christiana*, 35, 1864.

[179] J.J. Waterfall, F.P. Casey, R.N. Gutenkunst, K.S. Brown, C.R. Myers, P.W. Brouwer, V. Elser, and J.P. Sethna. Sloppy-model universality class and the vandermonde matrix. *Physical Review Letters*, 97, 2006.

[180] H. Wiley and P.M. Burke. Regulation of receptor tyrosine kinase signaling by endocytic trafficking. *Traffic*, 2(1):12, 2001.

[181] H.S. Wiley, S.Y. Shvartsman, and D.A. Lauffenburger. Computational modeling of the egf-receptor system: A paradigm for systems biology. *Trends in Cell Biology*, 13(1):43–50, Jan 2003.

[182] A.K. Wilkins, P.I. Barton, and B. Tidor. The Per2 negative feedback loop sets the period in the mammalian circadian clock mechanism. *PLoS Computational Biology*, 3(12):e242, 2007.

[183] A.K. Wilkins, B. Tidor, J. White, and P.I. Barton. Sensitivity Analysis for Oscillating Dynamical Systems. *SIAM Journal on Scientific Computing*, 31:2706, 2009.

[184] Y. Yarden. The egfr family and its ligands in human cancer. signalling mechanisms and therapeutic opportunities. *European Journal of Cancer*, 37 Suppl 4:S3–8, Sep 2001.

[185] Y. Yarden and M.X. Sliwkowski. Untangling the erbb signalling network. *Nature reviews. Molecular Cell Biology*, 2(2):127–37, Feb 2001.

[186] S. Yoon and R. Seger. The extracellular signal-regulated kinase: multiple substrates regulate diverse cellular functions. *Growth Factors*, 24(1):21–44, 2006.

[187] B.B.C. Youan. Chronopharmaceutics: gimmick or clinically relevant approach to drug delivery? *Journal of Controlled Release*, 98(3):337–353, 2004.

[188] D. Yu and M.C. Hung. Role of erb B2 in breast cancer chemosensitivity. *BioEssays News and Reviews in Molecular, Cellular and Developmental Biology*, 22(7), 2000.

[189] H. Yue, M. Brown, J. Knowles, H. Wang, D.S. Broomhead, and D.B. Kell. Insights into the behaviour of systems biology models from dynamic sensitivity and identifiability analysis: a case study of an NF-κB signalling pathway. *Molecular BioSystems*, 2(12):640–649, 2006.

[190] D.E. Zak, J. Stelling, and F.J. Doyle. Sensitivity analysis of oscillatory (bio) chemical systems. *Computers and Chemical Engineering*, 29(3):663–673, 2005.

[191] Y. Zhang, A. Wolf-Yadlin, P.L. Ross, D.J. Pappin, J. Rush, D.A. Lauffenburger, and F.M. White. Time-resolved mass spectrometry of tyrosine phosphorylation sites in the epidermal growth factor receptor signaling network reveals dynamic modules. *Molecular and Cellular Proteomics*, 4(9):1240–1250, 2005.

[192] J. Zhao, D. Ridgway, G. Broderick, A. Kovalenko, and M. Ellison. Extraction of elementary rate constants from global network analysis of e. coli central metabolism. *BMC Systems Biology*, 2(1):41, 2008.